UNIVERSITÄT BONN
Physikalisches Institut

Development and Characterization of a DEPFET Pixel Prototype System for the ILC Vertex Detector

von
Robert Kohrs

For the future TeV-scale linear collider ILC (International Linear Collider) a vertex detector of unprecedented performance is needed to fully exploit its physics potential. By incorporating a field effect transistor into a fully depleted sensor substrate the DEPFET (Depleted Field Effect Transistor) sensor combines radiation detection and in-pixel amplification. For the operation at a linear collider the excellent noise performance of DEPFET pixels allows building very thin detectors with a high spatial resolution and a low power consumption. With this thesis a prototype system consisting of a 64×128 pixels sensor, dedicated steering and readout ASICs and a data acquisition board has been developed and successfully operated in the laboratory and under realistic conditions in beam test environments at DESY and CERN. A DEPFET matrix has been successfully read out using the on-chip zero-suppression of the readout chip CURO 2. The results of the system characterization and beam test results are presented.

Post address:
Nussallee 12
53115 Bonn
Germany

BONN-IR-2008-14
Bonn University
September 2008
ISSN-0172-8741

Development and Characterization of a DEPFET Pixel Prototype System for the ILC Vertex Detector

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Robert Kohrs

aus

Düsseldorf

Bonn 2008

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Universität Bonn

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn http://hss.ulb.uni-bonn.de/diss_online elektronisch publiziert.

Referent:	Prof. Dr. N. Wermes
Korreferent:	Prof. Dr. K. Desch
Tag der Promotion:	29.8.2008

Abstract: For the future TeV-scale linear collider ILC (International Linear Collider) a vertex detector of unprecedented performance is needed to fully exploit its physics potential. By incorporating a field effect transistor into a fully depleted sensor substrate the DEPFET (Depleted Field Effect Transistor) sensor combines radiation detection and in-pixel amplification. For the operation at a linear collider the excellent noise performance of DEPFET pixels allows building very thin detectors with a high spatial resolution and a low power consumption. With this thesis a prototype system consisting of a 64×128 pixels sensor, dedicated steering and readout ASICs and a data acquisition board has been developed and successfully operated in the laboratory and under realistic conditions in beam test environments at DESY and CERN. A DEPFET matrix has been successfully read out using the on-chip zero-suppression of the readout chip CURO 2. The results of the system characterization and beam test results are presented.

Contents

Introduction

Substantial discoveries in high energy physics in the past few decades have revolutionized our understanding of the laws of nature at the most fundamental level. The Standard Model of particle physics describes three of the four known fundamental interactions between the elementary particles that make up all matter. Despite the great success of the Standard Model the missing integration of gravity and neutrino masses and 'aesthetic' reasons suggest a higher, more fundamental theory. Most of the more sophisticated approaches for physics beyond the Standard Model are extensions of the Standard Model, such as the Minimal Supersymmetric Standard Model (MSSM). The Standard Model of particle physics has been confirmed by experiments over the past fifty years. Currently one more particle is predicted by the Standard Model which is not experimentally observed yet, the Higgs boson. Therefore, the most important task of particle physics is the discovery of the Higgs boson - if it exists - and the comprehensive characterization of its properties.
Complementary to the Large Hadron Collider (LHC), which is currently being commissioned, a high precision collider is needed for this task. The planned International Linear Collider (ILC) is an electron-positron collider with center of mass energies of up to 1 TeV. To fully exploit the physics potential of the ILC a vertex detector of unprecedented performance is needed, including a spatial resolution below $5\mu m$, a low material budget of 0.1% of a radiation length per layer and a fast readout of almost a billion pixels within $50\mu s$. One technology option for the ILC vertex detector is the DEPFET sensor.

The DEPFET (DEPleted Field Effect Transistor) is a unique active pixel semiconductor detector. A first amplification stage in each pixel is integrated onto a high ohmic detector substrate. By means of sidewards depletion and an n-implant, a potential minimum for electrons is created underneath the transistor channel, acting as an 'internal gate' of the transistor. Due to the low capacitance of this internal gate an extremely low noise operation is possible.

This thesis discusses the development and the characterization of an ILC vertex detector prototype based on a DEPFET pixel sensor.

While the sensor is being developed and produced by the semiconductor laboratory of the Max-Planck-Institute (MPI) in Munich and the readout and steering ASICs[1] are designed in collaboration with the Lehrstuhl für Schaltungstechnik und Simulation (Circuit

[1]Abbr. for Application Specific Integrated Circuit.

Design) in Mannheim, in the Physical Institute of the University of Bonn the main focus is on the development of the system prototype. A detailed characterization of the system has been performed in the laboratory and in the beam test facilities at DESY[2] (Hamburg, Germany) and CERN[3] (Geneva, Switzerland).

This thesis is structured as follows:

- **1. The ILC Project** For a better understand the reasons for the demanding requirements, the first chapter briefly summarizes the physics goals of the ILC and the impact on the detector. The Large Detector Concept (LDC) is discussed in more detail as one of the proposed detector concepts. Finally, the requirements on the vertex detector and the baseline design will be presented.
- **2. DEPFET Active Pixel Sensors** After a brief introduction of the basics of radiation detection the DEPFET principle is discussed. The presented module concept for the ILC vertex detector is based on the proposal of the DEPFET collaboration.
- **3. The Prototype System** In this chapter the prototype system and the present components are described in detail.
- **4. System characterization** The basic characterization of the systems properties is presented, including a calibration of the full readout chain and a detailed timing and noise analysis.
- **5. Beam test studies** After a description of the experimental setup the event reconstruction methods are presented. The non-zero-suppressed data is analyzed with respect to signal and noise, charge sharing and spatial resolution. Different reconstruction methods like Center of Gravity and the so-called η-algorithm are discussed. The first operation of a high precision DEPFET beam telescope is discussed, followed by an approximation of the detection efficiency and purity.
- **6. Zero-suppressed readout** In this chapter the performance of the on-chip zero-suppression of the readout chip CURO 2 is discussed. The special considerations of a zero-suppressed event analysis are discussed and the results are presented. The observed issues of the readout chip are adressed.
- **7. Summary and Outlook** The main results of this thesis are summarized and potential improvements of the prototype system are given.

[2] Abbr. for Deutsches Elektronen SYnchrotron.

[3] Abbr. for Conseil Européen pour la Recherche Nucléaire, European Organization for Nuclear Research.

1. The ILC Project

1.1 Overview

There is a broad world-wide consensus among particle physicists that a linear electron-positron collider operating at a center of mass energy of 90 GeV up to 1 TeV be the next major collider facility for particle physics. The main reason for that is the high potential of complementary measurements to the Large Hadron Collider (LHC), currently being commissioned at CERN.
The LHC is a proton-proton collider with a center-of-mass energy of 14 TeV. It opens the door to a wide range of high energy physics and will allow answers to many of the most burning questions of particle physics, such as the origin of the electro-weak symmetry breaking. Furthermore, LHC will probably discover new physics beyond the standard model, like new symmetry laws in the TeV range or hints for a unification of all known forces and the structure of space-time.
However, the precision of measurements with a hadron collider is intrinsically limited due to the uncertainty of the initial energy of the interaction. The precise knowledge of the initial state of the interaction due to the point-like nature of the colliding particles is the main advantage of a lepton collider. Furthermore, all the center-of-mass energy is available in the primary collision. Other advantages are the tunable collision energy, the possibility of polarized beams (allowing for detailed analysis of the helicity structure of the process and sometimes significant suppression of backgrounds) and moderate backgrounds from Standard Model processes due to only electro-weakly interacting initial states leading to clean signatures.
There are several considerations leading to the decision for an e^+e^--collider using a linear accelerator. In principle, there are two approaches of accelerating particles for a collider experiment: the circular concept and the linear one. In case of the circular concept, particles are accelerated in a ring until they reach the desired energy and are then brought to collision. The beams are reusable, a refill of particles is usually required every few hours. When charged particles are being forced on a circular path synchrotron radiation is emitted, leading to a loss of collision energy. The energy loss ΔE for one circular revolution of a highly relativistic particle with rest mass m_0 and energy E given by

$$\Delta E = \frac{e^2}{3\varepsilon_0 (m_0 c^2)^4} \frac{E^4}{r} \tag{1.1}$$

(where e is the elementary charge, ε_0 the dielectric coefficient, c the speed of light and r the radius of the circular path), is acceptable for heavy particles like protons, but no longer feasible for electrons in the TeV range. The idea of using muons instead of electrons, which have a rest mass m_0 larger by a factor of more than 200, fails with the challenges of

an unstable particle with a decay time of 2 microseconds (for a feasibility study see [1]). In case of a linear collider, the particles start at two different locations, are accelerated on a straight track and are then brought to collision, as shown in fig. 1.1. Hence, the particle bunches are used for one collision only and cannot be reinserted. To stay within manageable dimensions in terms of geometry and cost, efficient accelerating cavities with much higher gradients than in the circular concept and an extreme beam focussing are required.

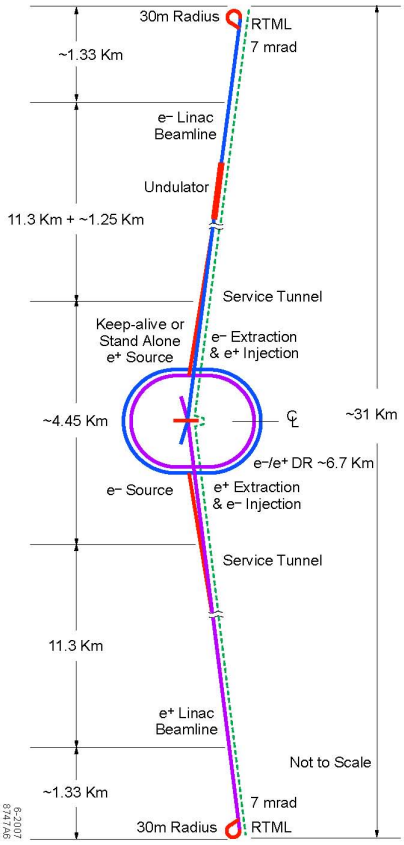


Figure 1.1: Schematic layout of the ILC complex for 500 GeV (cms) (from [28]).

The Luminosity is a measure of the expected rate dN/dt of a certain process characterized by its cross section σ:

$$\frac{dN}{dt} = \mathcal{L} \cdot \sigma \tag{1.2}$$

The total cross section σ_{total} of e^+e^- annihilation processes as plotted in fig. 1.2 shows a $1/s$ behavior [30]. To provide high statistics also for the high energy processes a luminosity in the order of $10^{34} cm^{-2} s^{-1}$ is envisaged at ILC.
In 2004 the decision for a superconducting accelerator technology was reached (as proposed by the former TELSA collaboration) and a world-wide project named International Linear Collider (ILC) was initiated. In 2005 a global design effort started. The goal is a detailed technical design of machine and detectors in 2008/09, with the baseline parameters of the machine listed below ([16]):

- Electron-positron collisions at $M_Z \leq \sqrt{s} \leq 500\,GeV$,

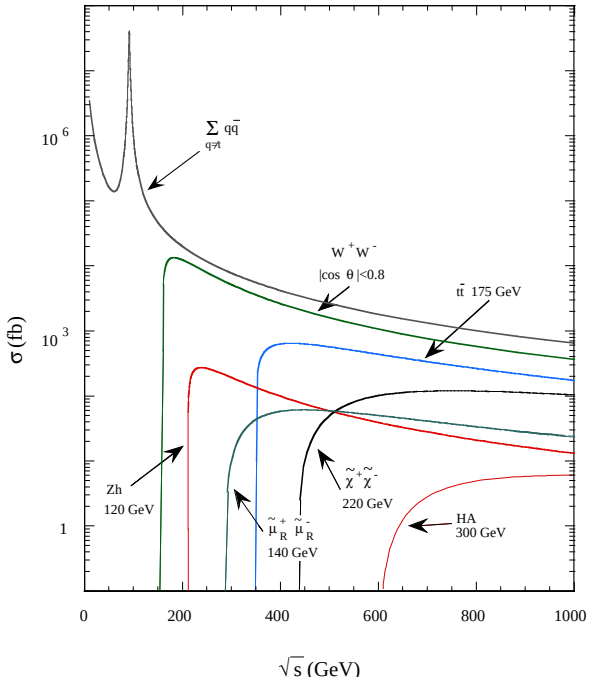


Figure 1.2: Cross sections of selected e^+e^- annihilation processes.

- Electron and positron polarization (electron polarization of at least 80%),
- Integrated luminosity of at least 500 fb^{-1} in the first four years,
- Upgradability to about 1 TeV with 500 fb^{-1} per year,
- Ability to scan in energy between 200 and 500 GeV (cms).

The next section summarizes the accelerator related boundary conditions like timing and bunching structure. In section 1.3 the Large Detector Concept is discussed as one of the proposed detector options. In the last section some general considerations of the vertex detector design are discussed.

1.2 Accelerator

The choice of the superconducting technology, as expressed by the International Technology Recommendation Panel (ITRP) in 2004, was an important step towards the realization of the ILC. The main advantage of superconducting cavities is the significantly lower surface resistance by a factor of 10^6 of niobium instead of copper structures operated at room temperature. This leads to a significant reduction of the power consumption. The overall transfer efficiency to the beam will be about 20 %. Besides those mainly economical reasons, the accelerator operation is simplified by the large cavity aperture and long bunch interval, the sensitivity to ground motion is reduced, and an increased beam current is possible.
The ILC is based on 1.3 GHz superconducting RF cavities operating at a gradient of 31.5 MV/m. The extremely high energy of the RF cavities requires a repetition rate

of 5 Hz with a beam pulse length of roughly 1 ms and a long pause of 199 ms between subsequent bunch trains (figure 1.3). Each bunch train consists of 2625 bunches of 2×10^{10} particles with a nominal spacing of 369.2 ns.
The beams are prepared in low energy damping rings that operate at 5 GeV and are 6.7 km in circumference. They are then accelerated in the main linacs which are ≈ 11 km per side. Finally, they are focused down to very small spot sizes at the collision point with a beam delivery system that is ≈ 2.2 km per side. To reach a peak luminosity of 2×10^{34} $cm^{-2}s^{-1}$, the collider requires ≈ 230 MW of electrical power. The beam itself has an elliptical shape of 640×5.7 nm^2 in order to achieve the high luminosity while keeping the beam related background at a tolerable level (see sec. 1.4). For comparison, this is 1/1000 of the size of the LEP beams in each dimension. A comprehensive discussion of the collider considerations can be found in [16].

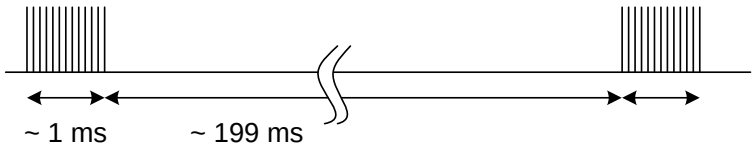


Fig. 1.3: Bunch structure of the ILC beam. A bunch train consists of 2625 bunches within roughly a millisecond, followed by a long pause of 199 ms.

1.3 Detector Concepts

Given the machine related constraints and the anticipated physics program some general considerations of the detector design arise. The main focus of the detector is precision.
At ILC, the absence of strong primary interactions leads to clean and completely reconstructible events. The relatively low interaction rate allows an event building without any hardware trigger, followed by a fully software based event selection, identification and analysis. Compared to the LHC detectors, radiation hardness is not such a critical issue at ILC. Thus many more detector technologies are available. In order to do physics with all final states, at the ILC a detector is needed which allows the precise measurement of all vertices with full hermiticity .
The physics at the ILC is dominated by the production of bosons (Z, W and potentially Higgs). All these bosons decay mainly (about 60 to 80%) to jets. Therefore, the reconstruction of the jets (or di-jets) becomes extremely important. Over the past few years the concept of particle flow has become widely accepted as the best method to reconstruct events at the ILC. The particle flow aims at reconstructing every particle in the event, both charged and neutral ones. A separation of particles is more important than the precise measurement of its parameters. This has in particular impact on the design of the calorimetric detectors, where excellent 3-D granularity becomes essential.
Track momentum and angular resolution are important for model independent measurements of the Higgs boson mass, ZH couplings and new particle masses in cascade processes (SUSY). Since flavor tagging is essential for many physics goals, in particular Higgs physics, the impact parameter resolution σ_b, i.e. the shortest distance of a particle

trajectory from the primary vertex, must be extremely precise. The impact parameter resolution of a detector depends on the point measurement precision, lever arms, mechanical stability and multiple scattering effects. The goal is a detector resolution of $\sigma_b(r\Phi, z) \leq 5\mu m \oplus \frac{10\mu m GeV/c}{psin^{3/2}\theta}$. Using the reconstructed secondary/tertiary vertex, the invariant mass of the tracks associated with decay is used to identify the jet flavor. Therefore a high granularity, high spatial resolution, thin layers and minimal distance to the interaction point are essential.
In the last few years, four different detector concept studies have emerged: the Global Large Detector (GLD), the Large Detector Concept (LDC), the Silicon Detector (SiD) and recently the Fourth Concept (4^{th}). Except of the latter, all of them follow the particle flow concept, but there are differences in focus and choice of technologies. The SiD concept foresees a 5T strong magnetic field and an all-silicon tracking system. LDC proposes a 4T field and relies on a large time projection chamber (TPC) supplemented by few layers of silicon vertex detector for tracking. The 3T field of the GLD requires an even larger calorimeter radius. The development of the detector concepts is still in progress, a Reference Design Report is currently in preparation.
The former TESLA detector proposal [37] is most closely related to the Large Detector Concept, and will be discussed in the following as the detector option for the ILC.

1.3.1 Large Detector Concept (LDC)

An overview of the detector concept is shown in fig. 1.4. Its key elements are a large volume tracker and highly granular electromagnetic and hadronic calorimeters, all inside a large volume magnetic field of up to 4 Tesla. The goal is an independent track reconstruction for the different detector components. The following sections give a brief overview of some detector components, details can be found in [24].

Tracking System A general layout of the proposed tracking system for LDC is shown in fig. 1.5. It comprises several distinct sub-detector systems.
A high precision pixel vertex detector surrounds the interaction point. It consists of five concentric layers with radii between 1.55 cm for the innermost layer and 6.00 cm for the outermost layer. This detector provides an excellent impact parameter accuracy, vertexing resolution and jet flavor tagging capability. The requirements on the vertex detector as well as the baseline design are discussed in detail in section 1.4.
Two concentric strip detector layers are arranged outside the vertex detector. Six disks, two of which are implemented as pixel detectors, cover the forward region (FTD, forward tracker).
The central tracker is a large volume, precision Time Projection Chamber (TPC) providing up to 200 points per track. The TPC is a gaseous chamber which is read out at the side via fine segmented Micro-pattern gas detectors (MPGD) at the endplate. As MPGD either Gas Electron Multiplier foils (GEMs) or the Micromegas technique are discussed. Compared to conventional wire chambers MPGDs have a better point resolution and two-track resolution as well as a better robustness in high backgrounds. The comparatively moderate single point resolution of a TPC of $r\varphi \approx 100\ \mu m$ is compensated by the large volume with fine granularity, allowing for an efficient and robust track finding even in

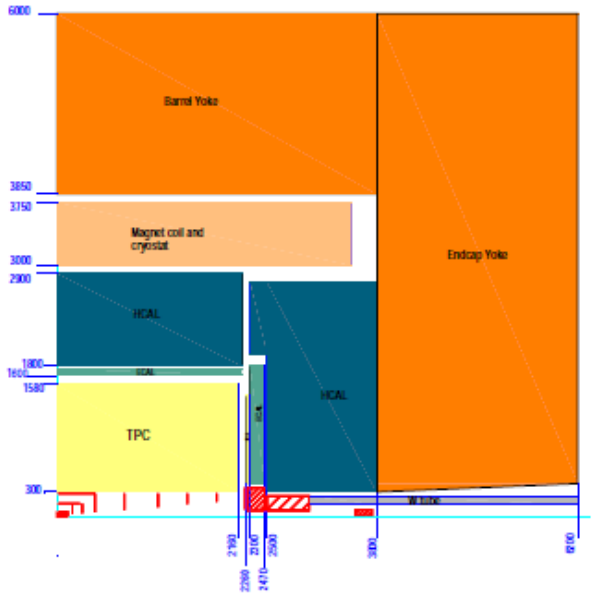


Fig. 1.4: Schematic overview of the LDC detector showing one quarter of the proposed design. All dimensions are in mm.

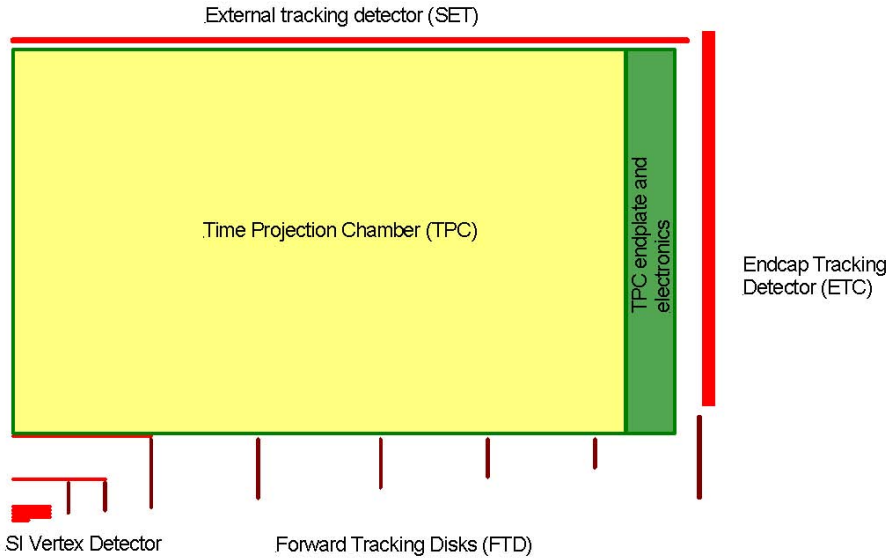


Fig. 1.5: General layout of the LDC tracking system. The complete tracking system is immersed in a 4T solenoidal magnetic field, aligned with the beam axis.

dense environments. Furthermore, the TPC offers particle identification via the specific energy loss (dE/dx) in the gas volume with an energy resolution of about 5 %. While the TPC plays a central role in the overall tracking concept, the final momentum resolution is reached by the combination of the TPC with the silicon tracking systems. The anticipated goal is $\delta(1/pt) \sim 5 \times 10^{-5}/\ GeV/\ c$, as defined by the $e^+e^- \rightarrow HZ \rightarrow H\mu\mu$ channel used for measuring the Higgs production rate and mass. This resolution is achievable with inner-silicon tracking and a TPC performance mentioned above.

Optionally the tracker can be upgraded by the inclusion of a layer of silicon strip detectors outside the TPC (SIT, silicon intermediate tagger), which provides additional calibration points in regions with little or no tracking coverage and helps to improve the linking between the TPC and ECAL.

Calorimetry The primary task of the calorimeter is to reconstruct the particle energy. However, within the concept of particle flow the role of the calorimeter is strengthened considerably. To fully exploit the physics potential of the machine precision measurements at the 3 %-level of the mass of pairs of hadronic jets are needed. The contribution of charged particles, which in average make up 65 % of a jet's energy, is measured with the tracking system. Neutral particles, i.e. photons and neutral hadrons, are measured with the calorimeter. The requirement of a jet energy resolution of $\frac{\delta E}{E} \simeq 0.30\ \frac{1}{\sqrt{E(GeV)}}$ implies that both electromagnetic and hadronic calorimetry are inside the coil to reduce inactive material in front of the calorimeters.

The electromagnetic calorimeter is divided into a cylindrical barrel and two end-caps. The general requirement about compactness, i.e. small Molière radius, has led to a sandwich calorimeter with a tungsten radiator (the possibility of lead is also studied). For the active part of the device, silicon PIN diodes seem perfect apart from their cost. Currently the calorimeter is segmented into readout cells of $5 \times 5\ mm^2$. The range of energies for photons and electrons suggests a thickness close to 24 radiation lengths for the ECAL. Current layouts foresee 20 absorbers with a thickness of 0.6 X_0 followed by another 9 absorbers of 1.2 X_0 thickness, resulting in a total thickness of the ECAL of only 18 cm.

The hadronic calorimeter in the particle flow concept must allow separating the energy depositions assigned to charged particles from those generated by neutral hadrons and thus eliminating the dominant part of hadronic jet energy fluctuations. The design of the HCAL in terms of technology and material is still under study. One option uses novel high gain scintillator cells ($SiPM$) with fine granularity and analog readout. A scintillator steel HCAL with $3 \times 3\ cm^2$ tiles was shown to meet the ILC performance goals in full simulation and particle flow reconstruction studies.

A second option is based on gaseous detectors and uses even finer granularity and digital readout. Iron (or Tungsten) is used as absorber material. The large number of readout channels allows reducing the resolution to a binary readout without compromising the the single particle energy resolution of hadrons.

In the very forward region a system of high precision extremely radiation hard calorimetric detectors are foreseen to measure luminosity and to monitor the quality of the collision (LumiCAL, BCAL, LHCAL). Details on the calorimeter studies can be found in [3].

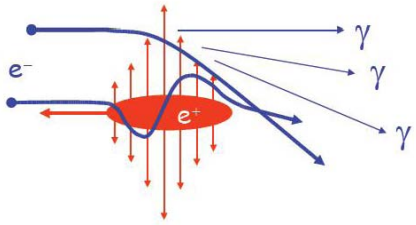


Fig. 1.6: Track of the electrons through the field of the opposing positron bunch.

1.4 Design considerations of the vertex detector

The anticipated physics programme requires a vertex detector with unprecedented tracking precision. Consequently, thin sensors with a low power consumption are needed, and the innermost layer has to be placed as close as possible to the interaction point. However, robustness against machine induced background and efficiency and purity in track reconstruction are as important as material budget and single point precision. Since a large background is expected, particularly close to the interaction point, multiple readouts during a bunch train become necessary for a robust track finding. In the following, the beam related pair production background is discussed as the dominant contribution to the background.

1.4.1 Beam related pair production background

Besides the number of particles per bunch N, the number of bunches per train N_B and the train repetition rate f_{rep}, the dimensions of the beam spots ($\sigma_{x,y}$) at the interaction point are the key parameters for a high luminosity. The luminosity $\mathcal{L}$ is increased by focusing the beam:

$$\mathcal{L} = \frac{N^2 N_B f_{rep}}{4\pi\sigma_x\sigma_y} \cdot H_P \quad \text{, where } H_P \text{ is the so-called pinch factor.} \tag{1.3}$$

A highly focused beam leads to the so-called *pinch effect*: during collision in an electron-positron collider the particles of each beam are accelerated towards the transverse center of the opposing bunch by its electric and magnetic forces. A further focussing of the beam is the result. The luminosity is increased by a factor of typically $H_P \approx 1.5 - 2$. This effect is in principal appreciated. However, bending (or even oscillations) of charged particles leads to the emission of radiation, the so-called *beamstrahlung*, whose spectrum is the same as of synchrotron radiation [35] (see fig. 1.6). According to [14] the energy loss due to beamstrahlung δ_{BS} is

$$\delta_{BS} \approx 0.86 \, \frac{e r_e^3}{2 m_0 c^2} \, \frac{E_{CM}}{\sigma_z} \left(\frac{N}{\sigma_x + \sigma_y} \right)^2 \tag{1.4}$$

The consequences are a wider energy spectrum of the colliding particles, an intense radiation collimated in a cone around the beam axis (which can be dangerous for the accelerator) and secondary e^+e^- pairs causing the dominant background. The pair creation

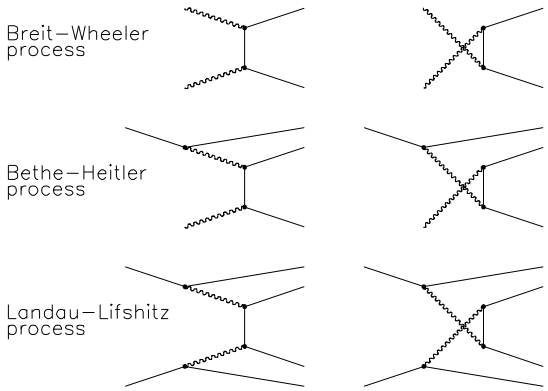


Fig. 1.7: The incoherent pair production processes.

is based on two processes, the coherent pair creation (CPC), where a photon interacts with the collective field of the oncoming bunch, and the incoherent pair creation (IPC), where a photon interacts with the field of an individual particle. While the CPC is almost completely suppressed for the ILC energy range, the IPC produces about 10^5 particles per bunch crossing with an average energy of some GeV. The most important production processes are called Breit-Wheeler, Bethe-Heitler and Landau-Lifshitz processes (see fig. 1.7).
Given the equations (1.3) and (1.4) the beamstrahlung can be minimized without affecting the luminosity by a flat (elliptical) shape of the beam ($\sigma_x \gg \sigma_y$). Different beam parameter sets are currently studied to find the optimum conditions for ILC.

1.4.2 Baseline of the vertex detector

The proposed baseline design of the ILC vertex detector is shown in figure 1.8. It consists of five layers of thin, high precision silicon pixel sensors.
The vertex detector is closest to the interaction point and therefore most exposed to the pair background. Fig. 1.9 shows the number of expected hits due to beamstrahlung as a function of the distance from the interaction point for a magnetic field of 4 T. Since the rate peaks at the interaction point and falls rapidly beyond, the vast majority of the particles can be prevented from hitting the detector by the combined effects of the strong, longitudinal magnetic field of the detector forcing the low energy particles on spiral tracks along the beam axis and a limited angular coverage.
Due to constraints of the machine collimation system the beam pipe will have a minimum radius of 14 mm [37]. Assuming the innermost layer of the vertex detector at that radius and a magnetic field of 4 T the expected background rate is approximately 0.05 hits per bunch crossing and mm^2. The occupancy of the pixel layers depends on the pixel pitch and the readout rate. A small pixel pitch is needed to keep the occupancy below a limit of around 1 % as required for robust track finding and to provide the required high spatial resolution. However, the demands on the readout speed are higher since more channels have to be read out. With a reasonable pixel pitch of 25×25 μm^2 and a typical cluster

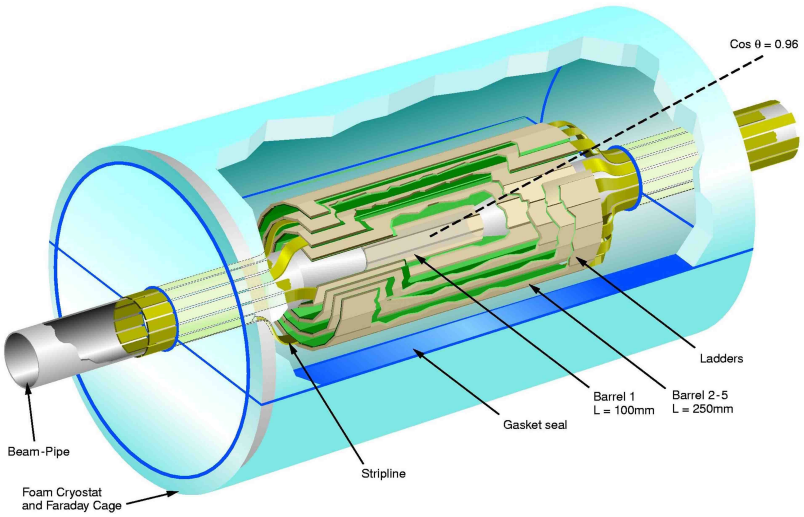


Fig. 1.8: General layout of the LDC micro vertex detector. An angular coverage of $|cos\theta| = 0.96$ is envisaged for the three innermost layers.

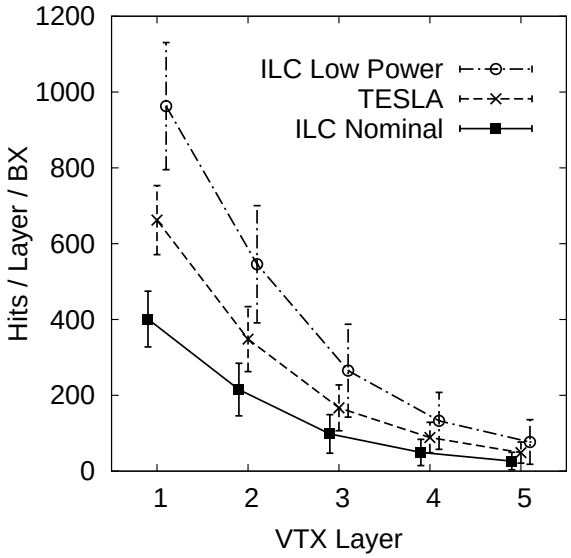


Fig. 1.9: Background hit rate vs. radial position, for different beam parameter sets (from [40]).

multiplicity of 2 pixels the number of expected hits per bunch crossing translates to an integrated occupancy of 15 % per bunch train. With the innermost layer having a length of 10 cm, 4000 pixel rows have to be read out roughly 20 times during a bunch train.
Since precision is one of the most important requirements for the vertex detector, the effects of multiple scattering have to be suppressed by reducing the material to a maximum of 0.1% of a radiation length X_0 per layer. This becomes possible, since an elaborate cooling is not a priori needed to minimize the radiation damage of the silicon sensors, as in the case of the LHC experiments. Instead, cooling gas flowing along the beam axis is foreseen, requiring the power dissipation of the vertex detector to be in the range of 10 Watts only. Further consequences of the low material budget are that the sensors have to be thinned to $\sim 50 - 100\mu m$. The use of a hybrid detector, where every pixel is directly connected to its own readout channel, is challenging particularly due to the small pixel pitch. Thus, the pixel array is read out sequentially, i.e. row-wise with a readout chip connected at the side of the matrix.
The requirements on the vertex detector, driven by the physics goals, are leading to a baseline design, which is almost independent of the technological realization. In chapter 2, a proposal for the ILC vertex detector based on DEPFET pixels will be presented. Except for the different strength of the magnetic fields having an impact on the beam related pair production background, the baseline design is similar for all three detector concepts. In conclusion the most important baseline performance parameters of the vertex detector as carried out by the Large Detector Concept studies are listed below (from [24]):

- Impact parameter resolution: $\sigma_b(r\Phi, z) \leq 5\mu m \oplus \frac{10\mu m GeV/c}{psin^{3/2}\theta}$
- Material budget: $X_0 \approx 0.1\%$ per layer.
- Power budget: In the order of 10 W for the whole vertex detector (allows gas cooling).
- Readout speed: A hit multiplicity of $\approx$ 0.05 hits per mm^2 and bunch at $\sqrt{s} = 500GeV$ must be tolerated with an occupancy of around 1%.
- Radiation tolerance: Electron fluxes of about 1.7×10^{12} per cm^2 and year, corresponding total ionizing dose of about 360 krad (for 10 years of operation) and a supplementary neutron equivalent flux of 8.5×10^{10} per cm^2 and year.

2. DEPFET Active Pixel Sensors

The principle of the **DEP**leted **F**ield **E**ffect **T**ransistor (DEPFET) was first introduced by Kemmer and Lutz in 1987 [20]. The DEPFET is a monolithic active pixel sensor device. The first amplification stage is integrated in each pixel. It is based on a fully depleted bulk, which yields an intrinsically fast and complete charge collection. With an extremely small input capacitance of the amplifying transistor this device is unique in terms of low noise and thus energy resolution making the DEPFET an attractive device in spectroscopic applications like autoradiography [18] or x-ray astronomy [17]. For the proposed application as vertex detector at the ILC, the high S/N ratio is exploited to build a thin sensor with a very precise spatial resolution and low power consumption.
In the first part of this chapter a brief and general introduction of semiconductor detectors is given. After the special operation principle of a DEPFET sensor is discussed in the second part, the conceptual design of the proposed DEPFET based vertex detector for the ILC, as reported in [37] and [4], is presented. The requirements and baseline features are discussed at this point. The details of the current chip generations and performance measurements can be found in the next chapters.

2.1 Basics of Radiation Detection

Based on fundamental interaction processes, radiation loses its energy partly or completely on its way through matter. Depending on the respective kind of radiation, this can be exploited for radiation detection in many different ways.
In the first part of this section, the interaction processes of radiation and matter are briefly summarized. Since the interaction processes are fundamentally different, charged particles and photons are treated separately. In the second part, the basic principle of semiconductor detectors is discussed. More on this subject can be found in [25].

2.1.1 Energy Loss of Charged Particles

The main processes of interaction between charged particles and matter are inelastic collisions with the atomic electrons of the material and elastic scattering from nuclei. Other processes, which are extremely rare by comparison, include the emission of Cherenkov radiation, nuclear reactions and bremsstrahlung. Heavy particles, i.e. everything heavier than electrons and positrons, predominantly lose their energy by inelastic collisions with atoms, causing an ionization or an excitation of the latter. In a single collision generally only a small fraction of the kinetic energy is transferred, but the number of collisions in normally dense matter is large enough for a substantial energy loss. Electrons and positrons in addition loose much of their energy by bremsstrahlung.

The mean energy loss of heavy, charged particles with medium and high energies in matter by ionization and excitation is given by the Bethe-Bloch formula [25]:

$$-\langle\frac{dE}{dx}\rangle = 2\pi N_a r_e^2 m_e c^2 \rho \frac{Z}{A}\frac{z^2}{\beta^2}\left[\ln\left(\frac{2m_e\gamma^2 v^2 W_{max}}{I^2}\right) - \beta^2 - \delta - 2\frac{C}{Z}\right] \tag{2.1}$$

with

r_e	:	classical electron radius
N_a	:	Avogadro constant
z	:	charge of incident particle in units of e
m_e	:	electron mass
ρ	:	density of absorbing material
Z	:	atomic number of absorbing material
A	:	atomic weight of absorbing material
β	:	v/c of the incident particle
γ	=	$1/\sqrt{(1-\beta^2)}$
I	:	mean excitation potential (for Si: $I = 172\,eV$)
W_{max}	:	maximum energy transfer
δ	:	density correction
C	:	shell correction

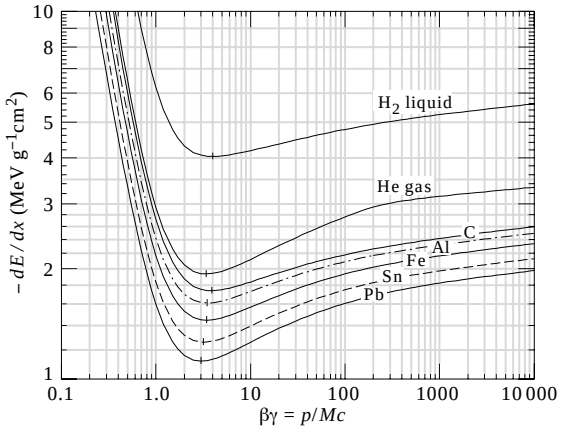


Fig. 2.1: Dependence of the mean energy loss on particle momentum for different absorber materials (from [12]).

The energy dependence of $\langle dE/dx\rangle$ is shown in figure 2.1, which plots the Bethe-Bloch formula as a function of the kinetic energy for several different absorber materials. For low energies, the curves are dominated by the $1/\beta^2$ term. At an energy which complies to 3-4 times the rest mass of the particle, the energy loss is minimal. Particles with energies larger than this are known as minimum ionizing particles (MIPs). With increasing energies, $1/\beta^2$ becomes almost constant and dE/dx rises again due to the logarithmic dependence

of equation 2.1.
Due to statistical variations of the processes, the energy loss of a particle crossing a relatively thick absorber, in which the amount of interactions is high, is of gaussian shape centered around this mean value. However, this is not true for thin absorbers as typically used in high energy physics experiments. Here, rare events with a large energy transfer to a single electron leading to secondary ionization (δ-electrons) add a long tail to the high energy side of the energy-loss probability distribution. The energy loss distribution for the case of a thin absorber was first calculated by Landau [23]. A typical Landau distribution is shown in fig. 2.2. Note that the mean energy loss no longer corresponds to the peak but is displaced to higher energies. The position of the peak defines the *most-probable* energy loss. In a 300 μm thick Si absorber the most probable energy loss is $\approx 84\ keV$, while the mean value is $\approx 120\ keV$.

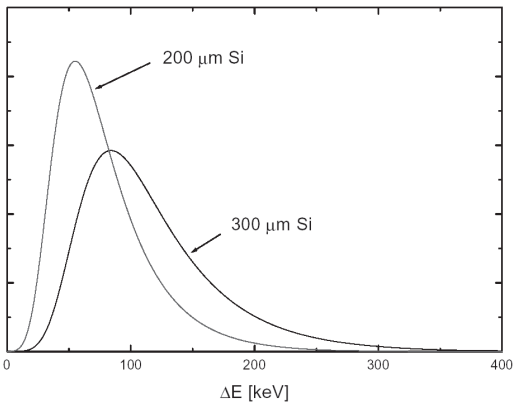


Figure 2.2: Shape of the Landau-distributions for two Si-detectors of different thickness. The scale of the vertical axis is arbitrary.

2.1.2 Interaction of Photons

The interaction processes of photons and matter are different from those of charged particles. The main interactions of X-rays and γ-rays in matter are:

- **Photoelectric effect:** the incident photon is absorbed by an atomic electron with the subsequent ejection of the electron from the atom. The energy of the outgoing electron $E = h\nu - \mathrm{BE}$ (BE the binding energy of the electron) is being transferred to the absorbing material by ionizing collisions as described above. The hole in the atomic shell left by the freed electron is filled by another atomic electron, giving rise to the emission of a characteristic X-ray or an Auger electron. For reasons of momentum conservation a nucleus is needed for the photoelectric effect.

- **Compton scattering:** the incident photon is scattered elastically on free or quasi free electrons[1]. The transferred energy depends on the energy of the photon and the scattering angle and is maximal for backscattering.

[1] If the photon energy is high with respect to the binding energy, this latter can be ignored and the electron can be considered as free.

- **Pair production:** the photon is transformed into an electron-positron pair. In order to conserve momentum, a third body, usually a nucleus is needed. Moreover, to create a pair, the photon must have at least an energy of 1.022 MeV.

As shown in figure 2.3 the probability of the respective interaction depends on the energy of the incident photon.

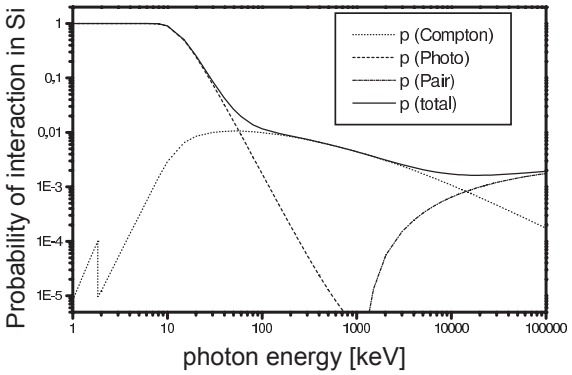


Figure 2.3: Probability of photon interactions in silicon as a function of the photon energy.

Since each interaction of a photon in matter leads to either absorption or to a deflection the intensity of the photon beam measured in the initial direction decreases exponentially:

$$I(z) = I_0 e^{-\mu z} \tag{2.2}$$

Here, z is the depth of the absorbing material and μ is the absorption coefficient, which is characteristic of the absorbing material and is directly related to the total interaction cross-section.

2.1.3 Semiconductor Detectors

Depending on the energy of the band gap, in semiconductors only few eV are needed to lift an electron from the valence to the conductance band and thus create an electron-hole pair. Including the energy spent on exciting lattice vibrations (phonons), the average energy needed to create an electron-hole pair in silicon at room temperature is 3.61 eV ($E_{gap} = 1.12\ eV$). This is very small compared to gaseous ionization detectors, where typically 20 eV is needed per electron-ion-pair, thus leading to a greatly improved energy resolution.
A basic semiconductor sensor for high energy physics applications is a silicon diode structure, as shown in figure 2.4. By applying a high voltage in reverse direction, the weakly doped (high resistivity) substrate is being fully depleted. Electron-hole-pairs generated by impinging particles are separated in the electrical field and drift to either side. The induced charge on the structured electrode is usually used as signal, which is amplified by an external charge-sensitive amplifier.
Assuming a typical thickness of a silicon detector of 300 μm, the most probable energy loss of a minimal ionizing particle is $\approx 84\ keV$. This translates into ≈ 23000 electron-hole pairs, which are generated along the path of the particle through the detector bulk. During the drift time of the charge carriers to the electrodes, diffusion in the lateral direction

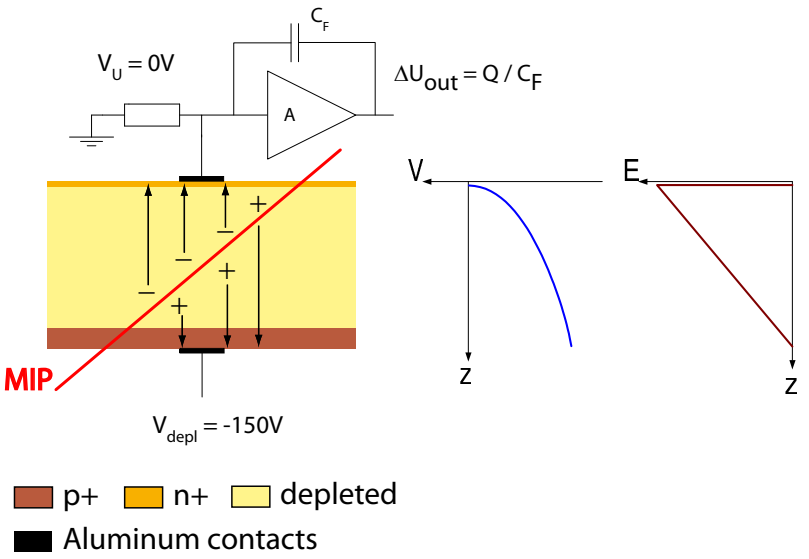


Fig. 2.4: The depleted pn diode used as particle detector.

leads to a widening of the charge cloud, as calculated in [26]. The observed charge distribution at the surface of the detector is in first order of gaussian shape[2]. In the case of a 450 μm Si detector studied in this thesis a charge cloud with a diameter of $d \approx 35.6\ \mu m$ ($\pm 3\sigma$) is expected for a bias voltage of $V_{bias} = -180\ V$.
The impact position of the particle can be measured by dividing the electrode on one or both sides of the detector into several smaller electrodes connected to their own readout channels. As shown in fig. 2.5, common designs include 1- and 2-dimensional strip detectors and 2-dimensional pixel detectors.

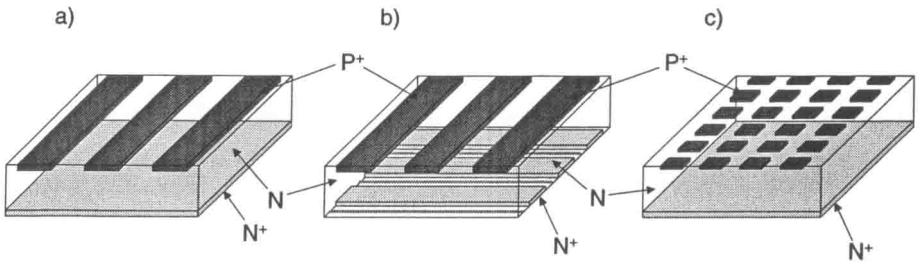


Fig. 2.5: Examples of position sensitive silicon detectors: single-sided (a) and double-sided (b) strip detectors and a pixel detector (c).

[2] The drift time and thus the diffusion of the charge carriers depends on the drift distance and the collecting field. Since charge carriers are generated all along the particle path through the bulk, the observed charge distribution is a superposition of many gaussians with different widths.

2.2 The DEPFET Principle

2.2.1 Sidewards Depletion

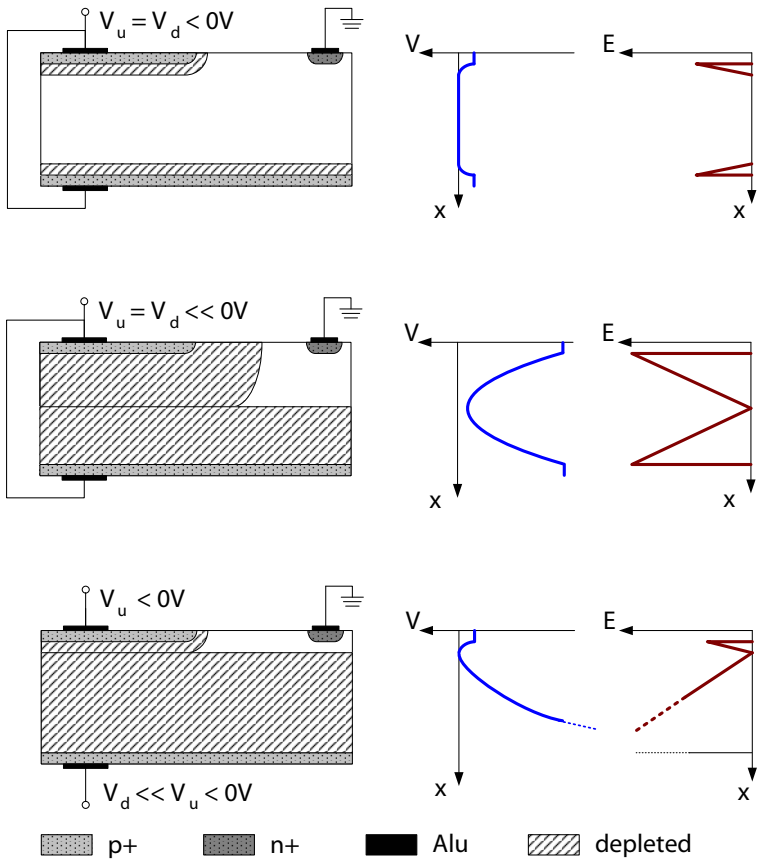


Fig. 2.6: The principle of sidewards depletion. Distribution of the resulting electrical field (right) and the potential (middle) in the detector substrate for different biasing conditions.

The principle of sidewards depletion as shown in figure 2.6 was proposed by Gatti and Rehak in 1984 [13]. In this concept p-implantations are processed on both sides of an n-type bulk. With the bulk contact at ground potential a reverse bias at the p-contacts leads to depletion regions at both pn-junctions.
The potential $\Phi(z)$ perpendicular to the substrate surface is described by the one-dimensional Poisson-equation:

$$\frac{\delta^2\Phi(z)}{\delta z^2} = -\frac{\rho}{\varepsilon_0\varepsilon_r}, \tag{2.3}$$

with the charge carrier concentration ρ. For different potentials at the front-side V_u and the backside V_d and a substrate of thickness d the Poisson-equation is solved by:

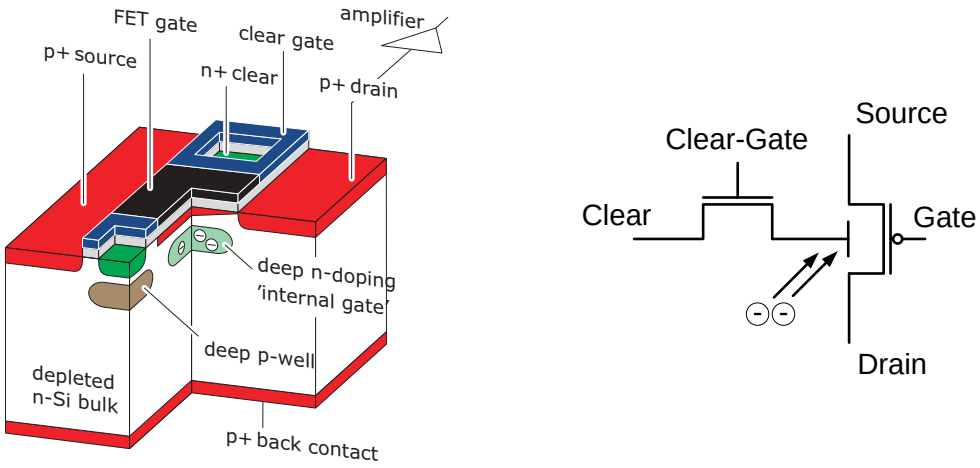
$$\Phi(z) = \frac{\rho}{2\varepsilon_0\varepsilon_r} z(d-z) + \frac{z}{d}(V_d - V_u) + V_u. \tag{2.4}$$

The position of the potential minimum z_{min} for electrons is given by:

$$z_{min} = \frac{d}{2} + \frac{\varepsilon_0 \varepsilon_r}{\rho d}(V_d - V_u). \tag{2.5}$$

Thus, as long as the bulk is fully depleted, the depth of the potential minimum is defined by choosing the potentials V_u and V_d.

2.2.2 Detector Principle



(a) Cross section of a linear DEPFET pixel.

(b) Equivalent circuit of a DEPFET pixel.

Fig. 2.7: Signal electrons are collected in a potential minimum underneath the external gate and modulate the transistor current. The accumulated charge is being removed from the internal gate via the clear structure.

A cross-section of a DEPFET pixel is shown in fig. 2.7. It is based on high resistivity n-type silicon substrate with an unstructured p^+ backside implantation. On the other side a p-channel FET is integrated in each pixel. An n^+ bulk contact is located outside the sensitive volume (not shown in fig. 2.7). By sidewards depletion a potential minimum for electrons is created in a typical depth of $\sim 1\ \mu m$ beneath the front side. Electron-hole pairs generated by impinging particles are separated in the electrical field. While the holes drift to the backside contact, the electrons are accumulated in the potential minimum. The lateral separation of the pixels is obtained by a supplementary, structured n^+-implantation (deep n), which in addition intensifies the potential minimum and shifts it further to the surface. This deep n implantation is located right underneath the conducting channel of the transistor and can be considered as an *internal gate* of the FET. Due to the capacitive coupling of the internal gate to the transistor channel, the transistor's conductance is controlled by the internal gate's potential, which in turn depends directly on the amount of collected charge carriers. Therefore, the change in transistor current δI_D is a direct measure of the change in charge δQ in the internal gate. The ratio of transistor current

δI_D to collected charge δQ is the device's amplification g_Q:

$$g_Q = \delta I_D / \delta Q = r \cdot \frac{g_m}{C_{ig}} \tag{2.6}$$

where g_m is the transconductance of the transistor, C_{ig} the capacity of the internal gate and r the fraction of the internal gate's capacity which couples to the transistor channel. The
The g_Q mainly depends on the gate geometry (W/L) of the transistor and the transistor current I_{DS}. Present devices show a g_Q of $\approx 300\ pA/e^-$.
Due to the small distance between internal gate and transistor channel stray capacitances are significantly suppressed. This yields an extremely low effective input capacitance of the amplifying transistor in the order of 10 fF and thus allows low noise measurements with an unprecedented energy resolution. ^{55}Fe-source measurements at room temperature using a linear DEPFET single pixel in ILC layout and a shaping time of 10 μs showed an rms noise of only $ENC = 1.6\ e^-$ [5].
The accumulated charge in the internal gate is isolated from the surrounding areas and remains there until it is being removed by a so-called clear process (see below). Furthermore, the charge collection is independent of the operating state of the external transistor. This allows an integration of the signal charge while the pixel is turned off and a readout on demand.
In contrast to diffusion based active pixel approaches like MAPS[3], the drift field in a fully depleted substrate offers a fast and complete charge collection. These properties lead to sensors with an intrinsically high signal to noise ratio, which is advantageous in many respects. First, the detection efficiency and purity are higher if it is easier to distinguish between signal and noise. Second, the analog pulse height information of the pixels can be used to improve the position reconstruction in case of charge sharing between neighboring pixels. The higher the signal to noise ratio, the more precise the reconstruction. Third, it can be exploited to build a thinner detector with a correspondingly smaller signal without compromising the detection efficiency.

Clear operation Besides the charge being generated by impinging particles also electrons originating from leakage current are collected in the internal gate. Since the collected charge changes the potential of the internal gate, it becomes less attractive for electrons. Therefore, the charge has to be removed regularly. Since the readout is non-destructive, this is done in a separate process by applying a positive voltage to a so-called *clear* contact, which is located adjacent to the transistor channel (see fig. 2.7).
A cross section of the region of the clear contact is shown in figure 2.8. The clear implantation is highly n-doped (n^+) to enforce the attraction for electrons and to provide an ohmic contact to the clear electrode. With a high positive voltage the clear contact becomes the most attractive region for electrons and the charge drift from the internal gate to the clear contact. During charge collection, however, charge drifting to the clear contact causes incomplete charge collection and must necessarily be avoided. Therefore, the n^+-clear region is embedded in a p-well, constituting a potential barrier for electrons.

[3] Monolithic Active Pixel Sensor

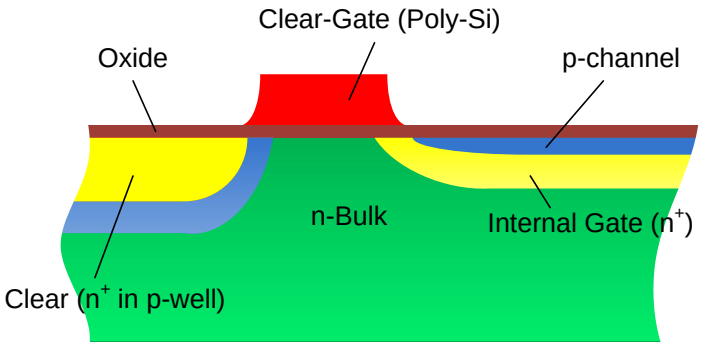


Fig. 2.8: Doping profile of the clear area.

Since charge remaining in the internal gate after clearing contributes to the total noise figure, a complete clear process is a crucial issue for the ILC application (see sec. 2.3.1). The competing demands of a complete charge collection during the integration phase and a complete removal of the charge after the readout process is one of the most delicate aspects of the pixel design. To ensure the clearing capabilities an additional MOS structure, the so-called *clear-gate*, has been introduced in the latest DEPFET generation. Depending on the design the clear-gates are either common over the whole matrix or connected row-wise for a clocked operation. By controlling the potential at the edge of source and drain region, the clear gate acts as insulation gate to prevent a parasitic edge current between source and drain. This renders a lateral channel isolation like LOCOS (local oxidation) or box channel isolation, as used in commercial CMOS processes, unnecessary.

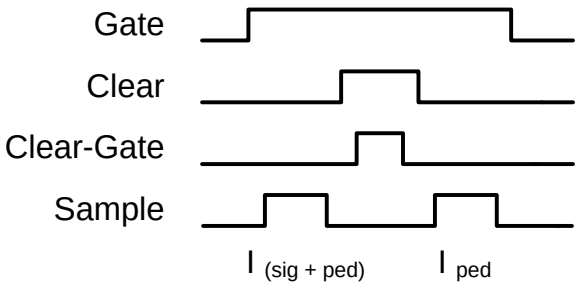


Fig. 2.9: Principal clear sequence.

As illustrated in fig. 2.8, the clear-gate covers the n-substrate, parts of the p-well and the tail of the deep-n implantation (internal gate). Given the different doping concentrations underneath the external gate the clear process is based on two mechanisms. First, a positive voltage at both clear and clear-gate leads to an inversion layer (n-channel) in the p-well connecting internal gate and clear. Electrons drift from the internal gate to the clear contact. Since the potential of the internal gate is not fixed by an external contact, it will get more and more positive. Depending on the actual potentials an equilibrium is reached in one of the following three cases:

1. The potential of the internal gate equals the clear potential.

2. No free charge carriers are left in the internal gate.

3. The conductivity of the inversion layer drops with increasing potential of the internal gate. Electrons could remain underneath the clear-gate electrode as long as the clear-gate is active. These electrons will flow back to the internal gate after lowering the clear-gate voltage.

In the latter case of an incomplete clear the second clear mechanism becomes important: a punch-through through the p-well potential barrier. For this mechanism the potential difference of clear and internal gate must be high enough. Lowering the clear-gate potential will ease this process. Finally, after successfully removing all electrons the clear potential can be lowered again. The principal clear sequence is shown in figure 2.9.

2.3 Module Concept for the ILC

2.3.1 Readout Concept

Operation of a DEPFET Matrix

The capability of a DEPFET to turn off the transistor without affecting the charge collection properties and the accumulation of the signal charge allow a row-wise readout scheme with a negligible dead-time. As shown in fig. 2.10, the transistor gate and clear contacts are connected row-wise to steering chips. The drain contacts are connected column-wise to a current readout chip. The source and clear-gate potentials are common over the whole matrix. By applying appropriate voltages at the gate contacts the transistors of only one matrix row are turned on at a time, while the transistors of all other rows are turned off. The drain currents are processed in parallel, allowing for a row-wise readout.

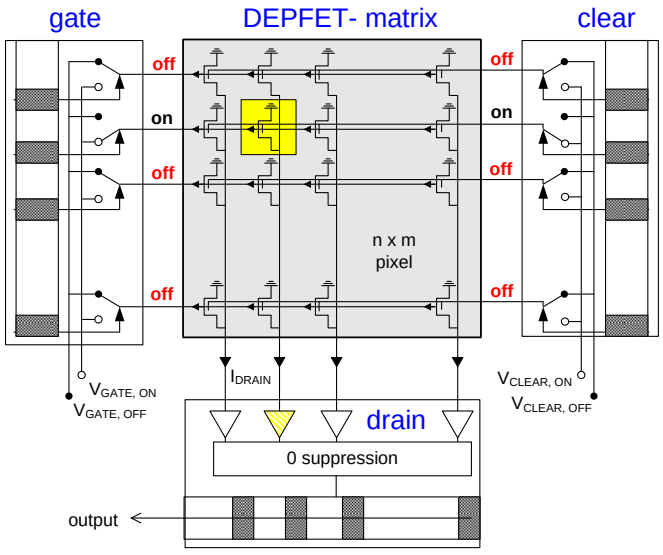


Figure 2.10: Operation principle of a DEPFET matrix. The transistor gates and clear contacts are connected row-wise to the steering chips, the drains are connected column-wise to the readout chip. By activating the transistors of only one row at a time while all other transistors are turned off, the matrix is read out row by row.

Readout scheme

The main task of the front-end chip is the triggerless, row-wise readout of the signal currents of the DEPFET matrix. As will be shown below the required frame rate at the ILC of $\sim$ 20 kHz translates into a row rate in the order of 10-20 MHz. Analog signals with a high signal to noise ratio are important for the precise spatial reconstruction of the hits. In addition, on-chip hit detection and zero-suppression are considered to be essential to cope with the vast amount of data. Hit detection is done by comparing the measured signal to a threshold value. Signals larger than the threshold are considered to be hits and are further processed, all other signals are discarded already on the front-end chip. Given these boundary conditions there are several options for a readout scheme.
In [39] it is shown that a voltage based readout using a source follower is far too slow. With a current based readout at the drain of the DEPFET transistor the settling time of the output signal is given by:

$$\tau = C_L \cdot R_{in} \tag{2.7}$$

with the load capacitance C_L and the input impedance of the succeeding readout circuit R_{in}, which can be kept sufficiently small using standard CMOS technology. Large matrices as needed for the ILC will have a high load capacitance in the order of 40 pF requiring a low input impedance of $R_{in} < 100\Omega$ to achieve a rise time τ in the order of a few nanoseconds. Therefore, a regulated cascode is used at the input of the readout chip, which provides a low input impedance and keeps the drain voltage at a constant potential.

The decision at which stage the current is being digitized has a big impact on the readout scheme and the design of the readout chip. An early digitization directly after the input cascode minimizes the analog deterioration of the signal. In turn, much faster ADCs and signal processing are required before zero-suppression.
Unavoidable process variations during the fabrication of sensors and readout chips lead to differences in noise behavior, offsets and gain between the channels, which affects the choice of the hit threshold. Thus, for a reliable on-chip hit discrimination a precise knowledge of the individual pixel pedestal current is needed. Assuming a pedestal current of $\approx 100\mu A$, a pedestal dispersion of 5% complies to $5\mu A$. This is already more than twice the expected signal of a MIP in $50\mu m$ silicon of $\approx 2\mu A$ (assuming an internal gain of the DEPFET of $g_Q = 0.5nA/e^-$). Thus, the pedestal of each pixel must either be stored on the chip or measured each time the pixel is read out, which would double the required ADC speed. Assuming, that 8 bits is sufficiently precise, the storage of 1000 pixels per readout channel for the innermost layer would require 1 $kByte$ of RAM per readout channel (and a factor of 2.5 more for the other layers).
The alternative to an early digitization is a fully analog zero-suppression. An individual threshold for each readout channel (i.e. per matrix column) is used in combination with a correlated double sampling (CDS). With CDS generally the output signal of a sensor is sampled twice within a very short interval. The first sample measures the signal itself. With the second sample the baseline (pedestal) is measured after a system reset. By subtracting both values all correlated contributions to the signal are canceled out. Spatial and temporal pedestal variations are suppressed resulting in a significant improvement of

the noise performance. In the present case the pedestal correction is done by comparing the analog DEPFET currents before and after a certain integration time.
The principle row-wise readout scheme is illustrated in fig. 2.11. At the end of one integration period, in which the other rows of the matrix are processed, the DEPFET current is sampled and temporarily stored in a current memory cell. After emptying the internal gates of that row via the clear process, the remaining pedestals are sampled and directly subtracted from the stored currents. The resulting current is the pedestal corrected signal, which originates from signal charge deposited by a traversing particle and leakage current (those two contributions are distinguishable by a common mode correction, see chapter 5 for further discussion):

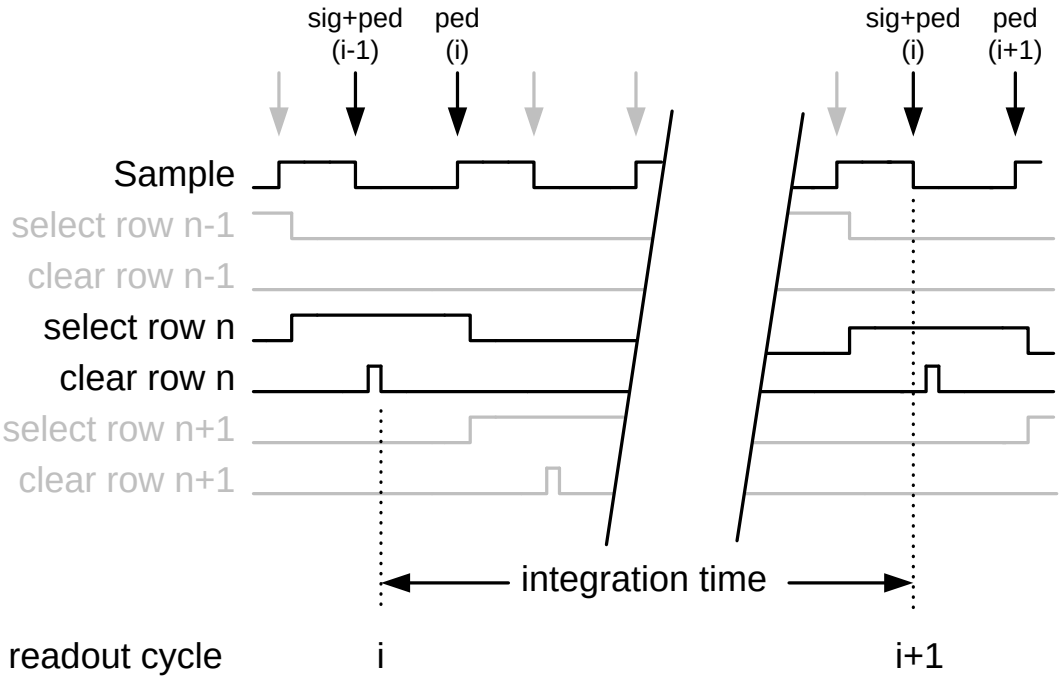


Fig. 2.11: Principle scheme of the foreseen row-wise readout of a DEPFET matrix. After selecting row n, signal plus pedestal current (i-1) are sampled and temporarily stored in the readout chip. With a clear pulse the signal electrons are removed from the internal gate. Then the resulting pedestal current (i) is sampled and subtracted from the buffered current, yielding the signal current plus the difference of pedestals (i-1) and (i). In case of a complete clear this difference is zero. Signals originating from impinging particles and leakage current can not be distinguished at this point.

$$I_{sig,i} = (I_{sig,i} + I_{ped,i}) - I_{ped,i+1}. \tag{2.8}$$

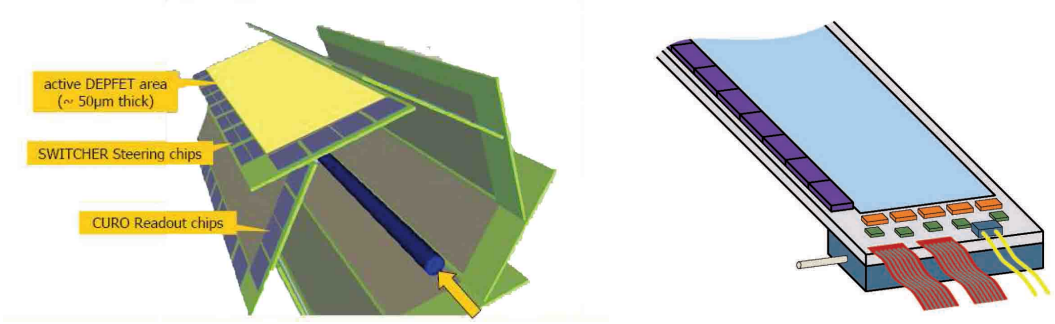
Since in this scheme the pedestals of two subsequent readout cycles are compared, a complete emptying of the internal gate (*complete clear*) is required to ensure a reproducible pedestal (see sec. 3.2). Otherwise, an additional noise contribution (reset noise) due to the remaining charge in the internal gate has to be considered.

Like all the components within the sensitive area close to the interaction point, also the readout chip must be radiation tolerant and material saving. The limited power dissipation is critical, as the usual trade-off is speed vs. power. In this context a power-down feature for the long pauses between the bunch trains is mandatory.

2.3.2 General Layout

The geometrical layout of the proposed vertex detector follows the baseline design as reported in [24]. It consists of five concentric layers of DEPFET sensors with a material budget of not more than 0.1 % of a radiation length per layer. Given the requirement of an angular coverage of $\eta = 0.96$ the outer layers, with a maximum radius of 60 mm, will have a length of 20 cm. Since the innermost layer will be located at a radius of only $r \approx 15\ mm$, a length of 10 cm is sufficient[4]. The default geometrical parameters are listed in table 2.1. Thus, the active area of one module will have a size of $\approx 13 \times 100\ mm^2$ and contain $\sim 512 \times 4096$ pixels with a size of $\sim 25 \times 25\ \mu m^2$. By deep anisotropic etching only the sensitive area of the silicon module is thinned to 50 μm, while the 300 μm thick edge regions serve as support frame (see 2.3.3). The rigidity of the frame and the tolerable deflection of the thinned part, such as bows under gravitational force, are currently being studied. Ideally, further support structures are rendered unnecessary. Most of the material, in particular the power dissipating parts like the readout electronics are situated on either short side of the ladder, outside the acceptance region, where a cooling pipe is arguable if necessary. Two sensor halves, each $\sim$ 2048 pixels long, are therefore processed in parallel. Due to the intrinsically low power consumption of the row-wise read out DEPFET array, where only a few transistors are activated for readout at the same time, cooling gas flowing along the beam axis is sufficient in the sensitive area.
Due to the different voltage levels of the steering functions (row selection and clear) an integration in one chip is challenging. Thus, two steering chips have to be placed on the matrix side. The steering chips are thinned to 50 μm and attached to the top side of the support frame along the long side by bump bonding technique. The signal traces from the steering chips to the sensor are integrated onto the support frame of the sensor module.
Keeping the occupancy at a reasonable level (≤ 1 %), a multiple readout during the bunch train becomes necessary. This requires a frame rate in the order of 20 kHz (see chapter 1). The speed requirements for the outer layers are more relaxed, since the background hit density is expected to be much lower at larger radii.
Note, that a row-wise readout imposes a much higher readout rate and - due to the lack of individual pixel calibration - the demands on the matrix homogeneity are high. A hybrid readout solution, where every pixel is connected to its own readout channel via bump bonds, would ease these two issues. However, apart from the challenge of the interconnection with an extremely small pixel pitch, the additional material of the ASICs and the interconnection inside the sensitive area is too high. On top, the vast power dissipation of in total $\sim 500\ M$ readout channels would necessitate elaborate cooling in the sensitive area, which would even increase the material budget. To suppress the effects

[4] Due to constraints on the design of the machine collimation system, it is necessary to set the beampipe radius at 14 mm [37].



(a) Innermost layer formed by 8 modules.

(b) One module. The readout chips are situated at the short edge of the sensitive area.

Fig. 2.12: Sketch of the proposed DEPFET vertex detector modules.

Tab. 2.1: Default geometrical parameters of the DEPFET based ILC vertex detector (from [4]).

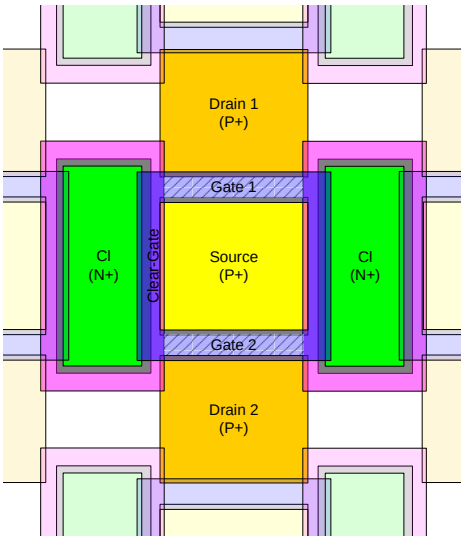
Layer	Number of ladders	Radius [mm]	Ladder length [mm]	width [mm]	readout time [μs]
1	8	15.5	100	13	50
2	8	26.0	2×125	22	250
3	12	38.0	2×125	22	250
4	16	49.0	2×125	22	250
5	20	60.0	2×125	22	250

of multiple scattering the material budget is limited to 0.1 % of a radiation length per layer though.

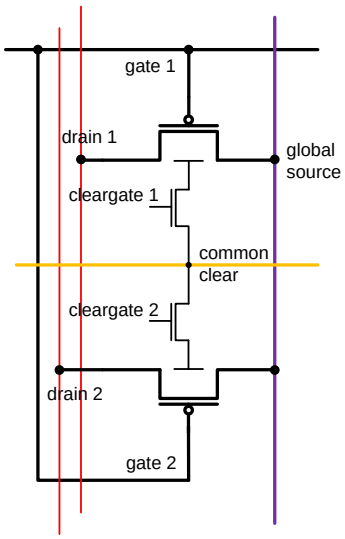
2.3.3 The ILC DEPFET Sensor

The DEPFET sensors are developed and fabricated in the MPI Semiconductor Laboratory in Munich. In order to meet the requirements for ILC the design of the DEPFET sensor has been optimized for large arrays of small pixels, fast readout and a complete clear.
Compared to JFETs [18] the chosen MOS technology yields higher pixel-to-pixel homogeneity due to self-alignment [15], allows smaller pixels and ensures process reliability. A further reduction of the pixel pitch is achieved by sharing the source and drain electrodes between the pixels, as illustrated in fig. 2.13. Both transistors of such a double cell are controlled by a common gate contact. Therefore, two pixel rows are readout in parallel via two drain lines per column, which doubles the effective row rate. The 4000 pixels of the innermost layer thus translate into 2000 rows of double pixels. By reading out the matrix at either end a frame rate of 20 kHz therefore translates into a row rate of 20 MHz. The clear contacts are located at either side of the pixel and are joined between neighboring

pixels in a row. Like the transistor gates the clear contacts are inter-connected row-wise, while the source potential is common over the whole matrix.
The current sensor generation (PXD-4) is discussed in more detail in the next chapters.



(a) One double pixel cell within the matrix array. The internal gates are located beneath the external gates (hatched areas) between source and drain implantations.



(b) Circuit of control lines.

Fig. 2.13: Double pixel cell of the linear DEPFET structures for the ILC. The two pixels share a common source implantation and the gates are connected allowing a simultaneous readout of two pixels via two separate drain lines.

Thinning Conventional thinning by backside grinding is not applicable with an electrically active backside implantation, while backside processing after thinning to a thickness of well below 100 μm is extremely difficult and cost intensive. Therefore, a thinning process has been developed at the MPI Semiconductor Laboratory in Munich.
The different processing steps are illustrated in figure 2.14. After bonding a backside-implanted, oxidized detector grade sensor wafer to a mechanical handling wafer by direct wafer bonding, the wafer sandwich is being mechanically thinned at the sensor side and processed as usual. The material of the handling wafer is then being removed only in the sensitive area of the sensor by deep anisotropic wet etching, which automatically stops at the silicon oxide interface of the wafer sandwich. This process allows to thin down the sensitive area in the center of a module to any desired thickness, while the outer areas remain at the initial thickness of typically 300 μm, thus serving as a support frame. A cavity pattern can be etched into that support frame for a further reduction of material (see 2.15). Such an 'all silicon ladder' needs no additional support material. This is in particular interesting, as mechanical stress due to different CTE of silicon and support structure is avoided. The unavoidable deflection from the ideal flatness of large-area, thinned devices is currently being studied. First results show a tolerable gravitational sag

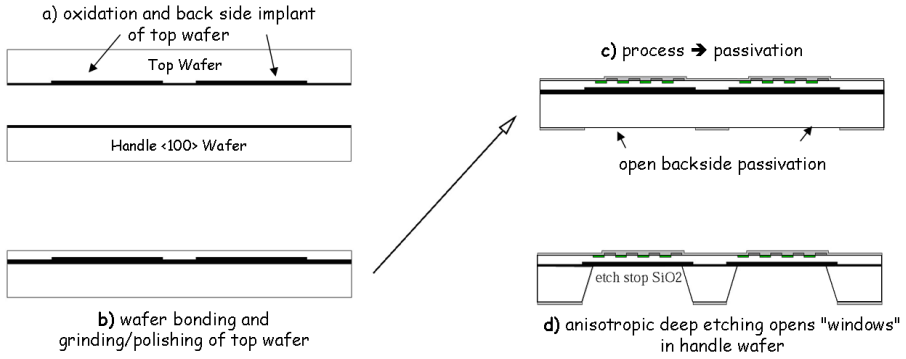


Fig. 2.14: Processing steps of the thinning technology for electrically active back side implants. After the oxidation of the top and handle wafer and the back side implantation for the sensor device (a) the two wafers are being connected by direct wafer bonding. The top wafer is then thinned and polished to the desired thickness (b). The processing of the top side is done on conventional equipment (c). The bulk of the handle wafer is removed by deep anisotropic wet etching. The etch process stops at the silicon oxide interface between the two wafers (d).

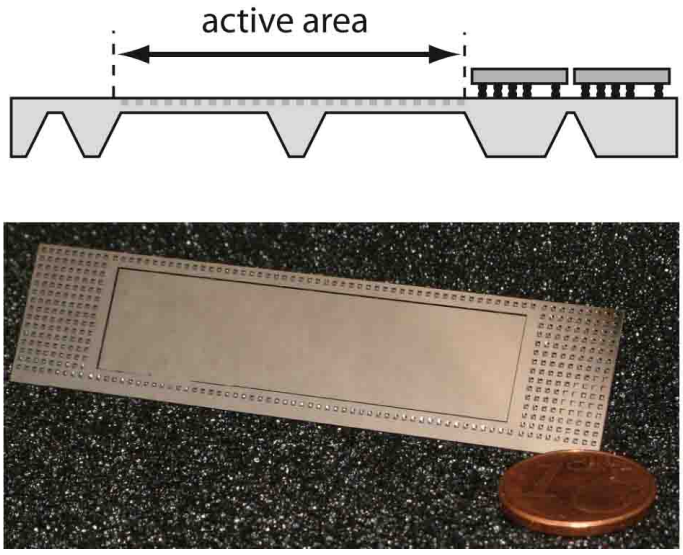


Figure 2.15: Cross section and photograph of a mockup module with thinned 'sensitive area' and a thicker silicon support frame. The readout and steering chips are directly bump-bonded onto the Si-support frame.

of about 20 μm over a length of 10 cm.
The feasibility has been demonstrated using diode structures. The deep anisotropic etching of the handle wafer does not deteriorate the characteristics of the PiN diodes and the handling of the etched wafers and diced thin chips with the integrated support frame was found to be safe and easy. The details of these measurements and the thinning process can be found in [22].

Radiation tolerance The expected total ionization dose (TID) at the innermost layers of the vertex detector is a few 100 $krad$ and an equivalent fluence of about 10^{11} 1 MeV-neutrons/cm^2 in a 5-years lifetime. This is about 3-4 orders of magnitude less compared to what is expected at LHC experiments like ATLAS, easing the requirements for the vertex detector considerably. The dominant source of the radiation at ILC is the pair background, but also synchrotron radiation and backscattered photons and neutrons contribute. Non-ionizing energy loss (NIEL) of the neutron background in principle causes bulk damage like an increase in leakage current, traps and change in the effective doping concentration. Since there is no charge transfer over long distances during the operation of a DEPFET (as in the case of CCD-type detectors), traps are of minor importance and a change in doping concentration is not expected at the doses relevant for ILC.
However, all MOS technologies, i.e. the DEPFET sensor and the steering and readout ASICs, are inherently susceptible to ionizing radiation. The dominant effect is a shift in threshold voltage to more negative values due to radiation induced charge accumulated in oxide and interface regions. Electron-hole pairs generated in the gate oxide due to ionizing radiation are to some extend separated in the existing field. Since the mobility of electrons in SiO_2 is 6 orders of magnitude larger than that of holes, holes may be trapped permanently in the transition region between silicon and oxide, while electrons diffuse through the barrier [33]. The accumulated positive charge carriers lead to an increase of the flatband voltage. Further effects are states in the interface between Si and SiO_2 resulting in an increased sub-threshold slope and possibly a higher 1/f noise of the transistors and a reduction of the transconductance g_m caused by a lower mobility of the charge carriers in the channel after irradiation.

To investigate the radiation tolerance of the MOS type DEPFET sensors, irradiation studies using ^{60}Co γ-radiation and hard X-rays (Molybdenum target) were performed with the current generation (PXD-4) using different biasing conditions. Note that this first iteration of DEPFET structures for an application at the ILC with its gate oxide of more than 200 μm thickness is not optimized for radiation tolerance. Depending on the operation state of transistor during irradiation the threshold voltage shift after annealing was found to be between -4 V and -6 V along with a moderate noise increase of 2 e^- ENC after almost 1 Mrad (Si) total ionization dose, demonstrating that the expected doses are not critical [10].
The threshold shift can easily be compensated by a continuous re-adjustment of the Gate switching level provided by the steering chip. The extra shot noise contribution due to the increased leakage current can be minimized significantly by moderate cooling of the device to values around 0° C.

2.3.4 Steering Chip

By activating only one row of the pixel array at a time and probing the current at the drains of the matrix, the sensor is read out row by row. The selection and the clearing of the active pixels is done by the steering chips, which toggle between two externally supplied voltages. Although the basic functionality can be provided by a simple voltage multiplexer, the chip must combine a voltage swing of $\approx 10\,V$ on load capacities of $> 20\,pF$ with rise and fall times in the order of a few ns, low power and radiation tolerance. By using an on-chip sequencer, the amount of external switching signals may be reduced. The requirement of radiation tolerance has impact on the choice of technology. Since commercial high-voltage processes come along with thick gate oxides, a standard deep sub-micron CMOS technology and radiation tolerant layout techniques are crucial.
The Switcher chips in their final layout have to be thin and small, the interconnection to the matrix will be done by fine-pitch bump-bonding. Several chips are daisy-chained in the case of large matrices.

3. The Prototype System

3.1 Overview

As a first step towards a full scale ILC ladder a prototype system has been developed. A schematic overview and a photograph are shown in figures 3.1 and 3.2, respectively. The system consists of two PCBs, a hybrid and a readout board. The hybrid hosts the highly sensitive parts like sensor matrix, steering chips and front-end chip. The DAQ (data acquisition) board, with an FPGA as central element, manages the synchronization of all components, the signal processing and data transfer to the PC. Besides a fast and low noise readout the design criteria of the prototype version were testability, flexibility and a slim housing for minimal distance of the planes in a beam test environment. During commissioning of the system the functionality has been proven and optimized with a dedicated CMOS test matrix [19]. In the following, details of the current generations of the Sensor and the ASICs are discussed. The description of the hardware is done on a rather functional level.

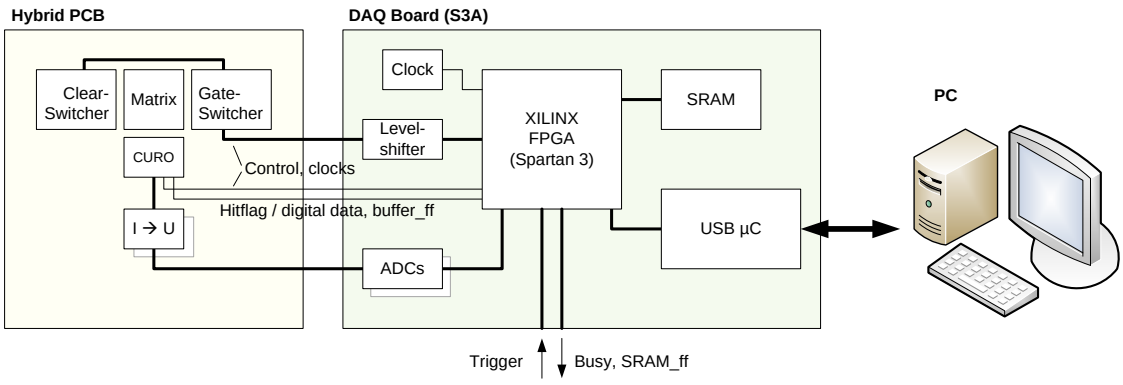


Fig. 3.1: Schematic of the S3A prototype module consisting of hybrid PCB and DAQ PCB. Only selected connections are shown.

3.2 The DEPFET Sensor (PXD-4)

With the first production of ILC type DEPFETs (PXD-4), many different matrices with up to 64×128 pixels have been fabricated on unthinned ($450\ \mu m$) wafers, featuring pixel sizes down to $24 \times 33\ \mu m^2$. Since the focus of former DEPFET designs was on spectroscopic devices, many layout modifications were necessary. To reach small pixel sizes the

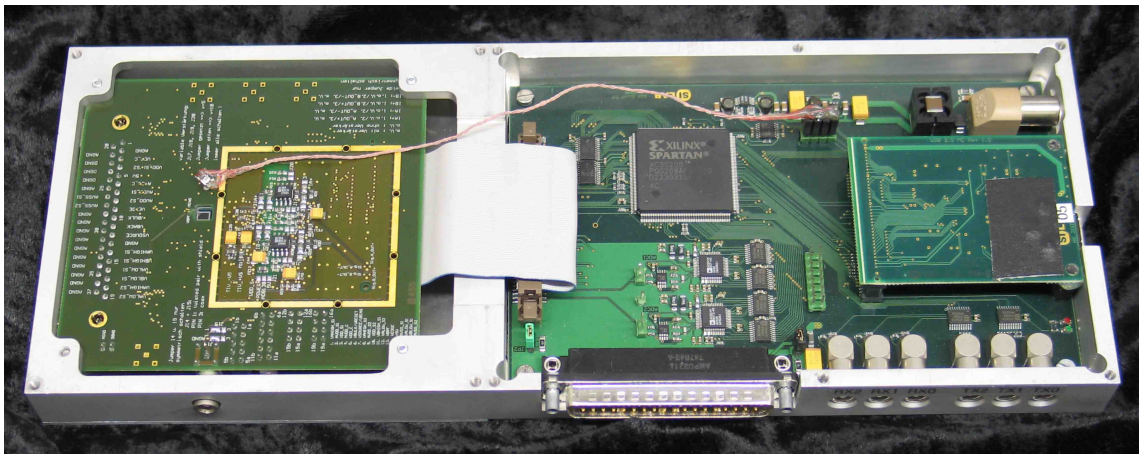


Fig. 3.2: Photograph of the DAQ board (right) and the backside of the hybrid (left) inside the aluminum frame, connected by a ribbon cable. The cut-out for the matrix is visible in the center of the hybrid. A temperature sensor is attached next to the matrix and connected to the I2C connector of the DAQ board.

technology has been changed from JFETs to MOSFETs and a double pixel structure has been introduced, as mentioned in section 2. The implementation of four conducting layers (two polysilicon and two metal layers with a low resistivity) addresses the requirements of a fast readout of large DEPFET arrays. Special attention has been payed to control the potentials in the region between internal gate and clear implantation to ensure a complete clear on one hand and to prevent charge loss on the other hand. For this reason, an additional MOS structure, the so-called clear-gate, has been introduced in the latest DEPFET generation (see chapter 2.2.2 for details on the clear mechanism).
In principle, the clear-gate potential can be pulsed, as for the clear contact. That means, that different voltages are possible during the clear process and during the charge collection. However, a complete clear with a static potential of the clear-gate would yield several advantages for the application at the ILC: no additional strobe of the clear-gate saves time and the steering chip. Furthermore, the simplified routing of the pixel array allows smaller pixels. With a static clear-gate, in turn, it is even harder to achieve a complete clear in combination with a complete charge collection. Choosing the potential too low may lead to an incomplete clear. If the potential is chosen too high, parts of the signal charge are collected in the substrate underneath the clear-gate instead of the internal gate (charge collection inefficiency).

Another design feature of the PXD-4 generation is an unstructured, deep high energy (high-E) n-implantation in a depth of $> 1\ \mu m$, which was realized for some structures. Due to this implantation, the potential minimum for electrons is generally shifted deeper into the substrate. The reduced capacitive coupling to the clear-gate eases the punch-through to the n-region beneath the clear-gate and the internal gate. This is in particular interesting for structures with static (common) clear-gate, since the whole clear process is done by punch-through (the n-channel is only effective at the surface, the punch-through is also effective in the depth). Furthermore, due to the slightly deeper internal gate ($\approx$ $1.2\ \mu m$ instead of the standard $\approx 0.65\ \mu m$), the signal electrons have less interaction with

Clear-gate	Pixel sizes	Internal notation
clocked	$36.0 \times 28.5 \mu m^2$	
common	$33.0 \times 23.75 \mu m^2$	'*rec small*'.
common	$36.0 \times 22.0 \mu m^2$	'*rec small A*'.

Tab. 3.1: Selected designs of the current DEPFET production PXD-4. All matrices feature 64×128 pixels.

the clear-interfaces during charge collection, which prevents charge loss. The drawback of a deeper internal gate is a reduced coupling to the transistor channel. Thus, the internal amplification is expected to be lower compared to a device without high-E implantation. The clear efficiency with respect to different clear and clear-gate voltages has been studied for various design options with and without high-E in detail and is reported in [34] and [2]. It has also been shown, that a fast clear of $< 10\ ns$ is possible using a static clear-gate with high-E implantation. Further details on the technology used in the current PXD-4 run is given in [11].
Besides many smaller test structures, 64×128 pixel matrices have been fabricated with different designs. A table of the most important ones and the respective pixel sizes are listed in table 3.1. All layouts are available with and without high-E implantation.

3.3 Steering Chip SWITCHER 2

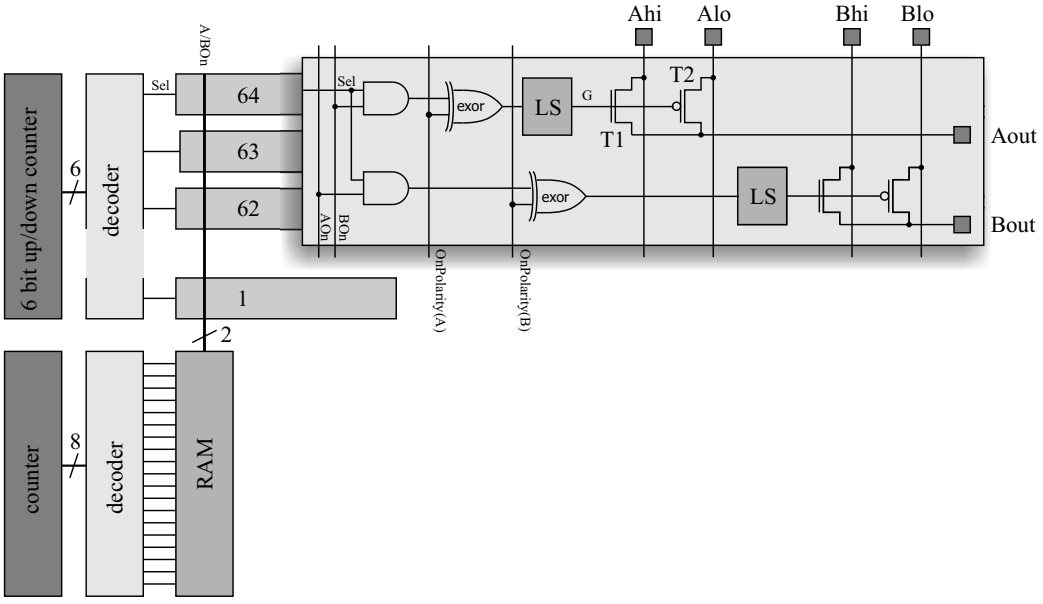


Fig. 3.3: Schematic overview of the SWITCHER 2 chip (from [29]).

The versatile steering chip SWITCHER 2 is fabricated using a 0.8 μm HV technology. The architecture of the chip is shown in figure 3.3. It is capable of switching the 2×64

output channels within a range of 25 V between its analog supply voltages (AVSS, AVDD). The digital ground level as well as the polarity and the levels of the switching voltages are adjustable. The high voltages were desired in this version of the design to allow all possible DEPFET device studies. Lower clear voltages as envisaged for the next DEPFET generations will allow the use of thinner gate oxides of the switching transistors with a better radiation tolerance. The output resistance of the pass transistors of typically 230 Ω for the pMOS (rising edge) and 150 Ω for the nMOS (falling edge) allows switching 20 V_{pp} with 35 MHz. The switching sequence may either be controlled by an internal sequencer or by externally supplied control strobes. Several SWITCHER chips can be daisy chained by I/O signals at the top and the bottom of the chip. A detailed description can be found in [29].

3.4 Readout Chip CURO 2

The current version of the readout chip is called CURO 2. It was designed in 2003 [39] and fabricated using a 0.25 μm deep sub micron process, which is intrinsically radiation tolerant. In the following the operation principle and the major building blocks of the CURO concept are discussed. Performance measurements can be found in the next chapter. Additional information and measurements can be found in [39].
The CURO concept foresees a triggerless operation with two continuous clocks, one for the input stage (analog part) synchronous to the row clock (Write-Clock) and a faster one for hit processing behind the comparator (Scan-Clock). Both clock regimes are separated by a small mixed signal FIFO, which derandomizes the hit rate fluctuations to a constant rate. The architecture of the CURO concept is illustrated in fig. 3.4. For simplicity reasons only one of the 128 channels is shown. Note, that the analog FIFO actually implemented in CURO 2 has only a depth of one, which means, that the derandomizing effect of the FIFO is missing. This has a large impact on the continuous operation; in particular the two clocks have to be pulsed in a certain scheme rather than being clocked continuously. The basic idea of a continuous signal processing is discussed in the following by means of fig. 2.11. The *Write-Clock* (*WrCLK*) is synchronized with the row clock, i.e. with each *WrCLK* cycle the next row of the matrix is being selected.

1. The current $I_{sig,i} + I_{ped,i}$ is sampled on the high level of the *WrCLK* and stored in the first memory cell following the cascode (1) with the falling edge.

2. By emptying the internal gates of the row the current of the matrix is reduced to the pedestal only. Since signal and pedestal current are still provided by the memory cell, the pedestal current is automatically subtracted at the input node of the memory cell according to Kirchhoff's law. This is done during the low level of the *WrCLK* and the value is sampled at the rising edge. To ease the timing of the succeeding processes the resulting signal current is stored alternately in two current buffer cells (e.g. A).

3. With the next *WrCLK* cycle the current stored in memory cell A is stored in the analog FIFO. At the input the next row of the matrix is already being processed as described above, which is ignored in the following.

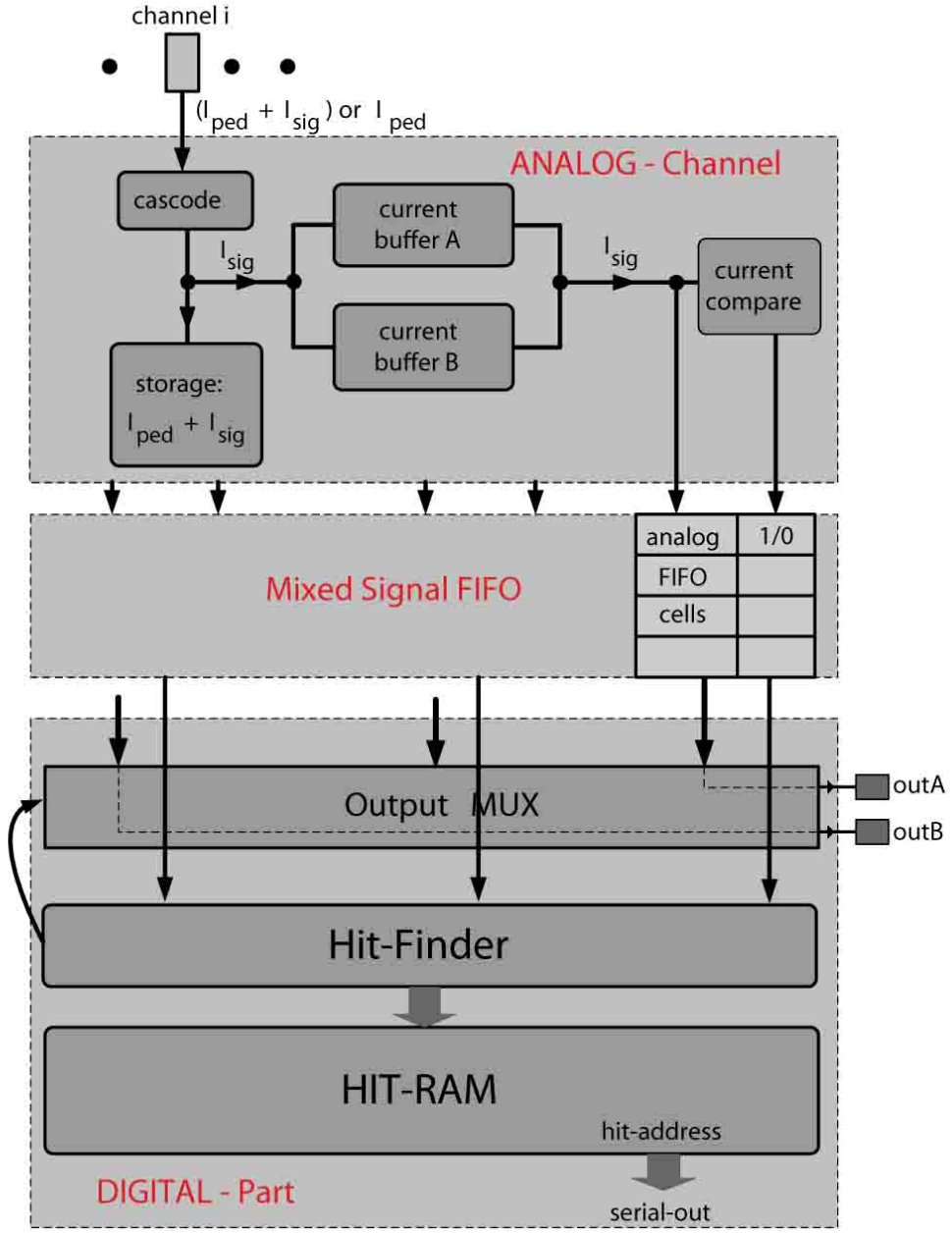


Fig. 3.4: Schematic overview of the CURO concept. In the present version CURO 2 the analog part of the mixed signal FIFO has a depth of one only, which limits the continuous operation of the chip.

4. The current, which is still stored in memory cell A, is compared with a threshold on the third *WrCLK* cycle. The binary result is stored in a digital FIFO.

5. The binary hit information is processed by a hit finder. With the first Scan-Clock (*ScCLK*) cycle the hitscanner is loaded. If hits are found, the first two analog values are multiplexed to either analog output instantaneously, while the hit addresses (row and column) are stored in a hit buffer for later readout (e.g. in the bunch pause). The detection of a hit is indicated by a so-called *hit-flag*.

6. With each *ScCLK* cycle up to two hits are processed in the same way. If no further hit is found the scanner resets the hit-flag. The hit pattern of the next row can only be loaded from the FIFO into the scanner if the hit-flag is low.

Analog FIFO The present version of CURO lacks an analog FIFO. A memory cell per channel is used to buffer the analog information, but the derandomizing effect of the FIFO is missing. Since the next row can therefore not be selected until the present one is fully processed a continuous data taking in the scheme described above is not possible with this version of CURO. Therefore, three *WrCLK* cycles are needed per row, during which the analog data stored in the buffer is protected from being overwritten. This *freezing* of the buffer is done by resetting the so-called *buffer_en-flag*.

Current memory cell The current memory cell is a central element of the concept. A careful design is essential to ensure a high bandwidth, accuracy and low noise at the same time. A detailed discussion of the design considerations can be found in [39]. To combine the competing demands as much as possible, the concept of a double cell has been chosen, where the first stage has a high bandwidth and thus a high noise while the second stage is slow but accurate (low bandwidth, low noise). Furthermore, the effects of charge injection are reduced and the linearity is improved. The simplified schematic of such a double cell is illustrated in fig. 3.5. The sampling process of one cell is divided into three phases:

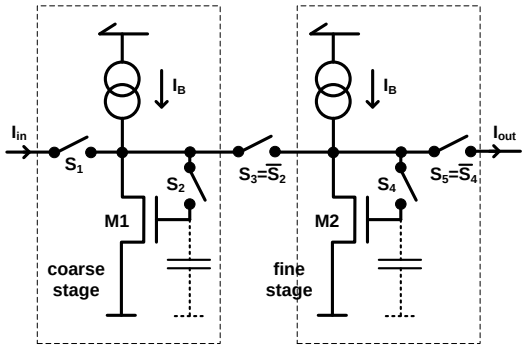


Figure 3.5: Basic principle of the double stage current memory cell (see text).

1. In the initial state the switches S1 and S2 are closed and S3 is open. The gate capacitance of transistor M1 is charged until the device provides the combined input and bias current ($I_{M1} = I_{in} + I_B$).

2. S2 is opened. The gate voltage and hence the transistor current ideally remain unchanged.

3. After sampling, S1 is opened and S3 is closed. The current through M1 must still be $I_{M1} = I_{in} + I_B$ due to the gate potential of M1. The switched off I_{in} must therefore be delivered by the output node to satisfy Kirchhoff's current law.

The timing of the different phases is derived from the *WrCLK*. The duration of the coarse sampling is adjustable in the order of nanoseconds by a deskewing logic. However, the maximum duration of the coarse stage is only 15 ns, independently of the duration of the fine sampling. Since the minimum clear duration with the present system is 20 ns, the coarse stage is active before or during the clear. Although the bandwidth of the double cell is designed to operate at 50 MHz, the achieved readout speed is therefore lower than required. This is discussed in chapter 4.4.

Current comparator The comparator generates a binary hit signal, depending on the signal amplitude. Signals are identified as hits, if they are above a certain threshold. The operation scheme is illustrated in fig. 3.6. The threshold current is subtracted from the signal current, coming from the buffer cell. The resulting current, which is either positive in case the signal is larger than the threshold or negative in the other case, charges or discharges the gate capacitances of the transistors M1 and M2. The output node between the transistors is either tied to ground or to V_{CC}. This digital information is stored in a flip-flop for further processing.

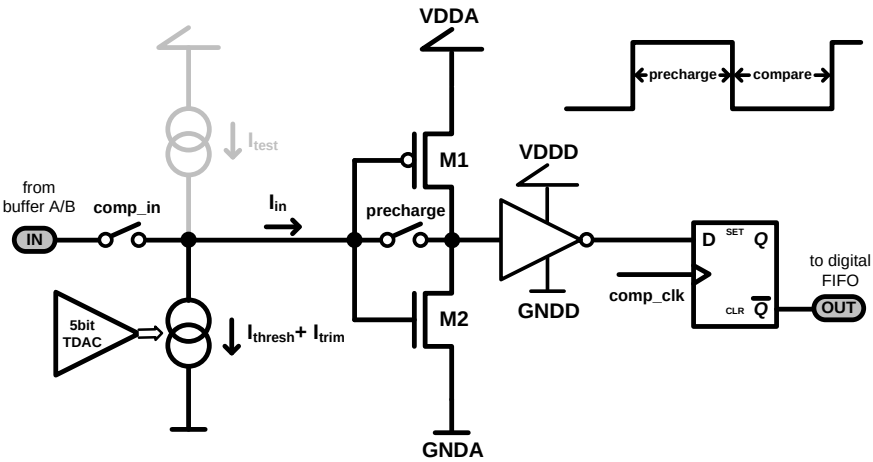


Fig. 3.6: Schematic of the comparator stage.

The input of the inverter (M1, M2) is pre-charged to the trip-point by a positive feedback in the first half of the clock cycle to increase the speed for inputs close to the threshold and to provide identical conditions for all signals. During the pre-charge phase of the circuit the inverter stage consumes maximum power, which is up to 75 mW for the whole chip. Since this high current is pulsed with the high *WrCLK* frequency, also voltage ripples

on the supply rail will occur. This is further discussed in chapter 5. To compensate for the unavoidable threshold dispersion the threshold can be tuned channel-wise by a 5-bit DAC. The tuning range is set globally by an 8-bit DAC ($I_{trim,\,LSB}$).

Hitscanner The mean processing time of the zero-suppression per row is given by the row rate, which in the current baseline is 20 MHz. Since the expected occupancy of 1.7 hits in a 128 pixel row is relatively low, a design with a parallel tree structure was chosen for hit identification. The principle is illustrated in fig. 3.7. In a first step, the existence of at least one hit is indicated by the hit-flag, which is realized as a logical OR in the out direction. Then, the hit addresses are traced back via the back propagating feature of the leaves, which includes a priority logic for the case of more than two hits. Since only logical gates are involved in both steps, the computing time is negligible. This design allows the identification of up to two hits per *ScCLK* cycle. After identification the hit addresses are binary coded and stored in the hit-buffer. The corresponding hits are reset in the input register in the next clock cycle and the scan starts with a reduced pattern. If the pattern is emptied completely the next pattern is loaded.
Since the minimum processing time per row is one *ScCLK* cycle in the case of no hit and linearly scales with the amount of detected hits, the frequency of the *ScCLK* must be higher than the row rate. However, given that most of the rows contain no hit, the future implementation of a small derandomizing FIFO will relax this issue. A simulation demonstrates that a FIFO depths of four and a *ScCLK* frequency of 40 MHz are sufficient for an efficient derandomization.

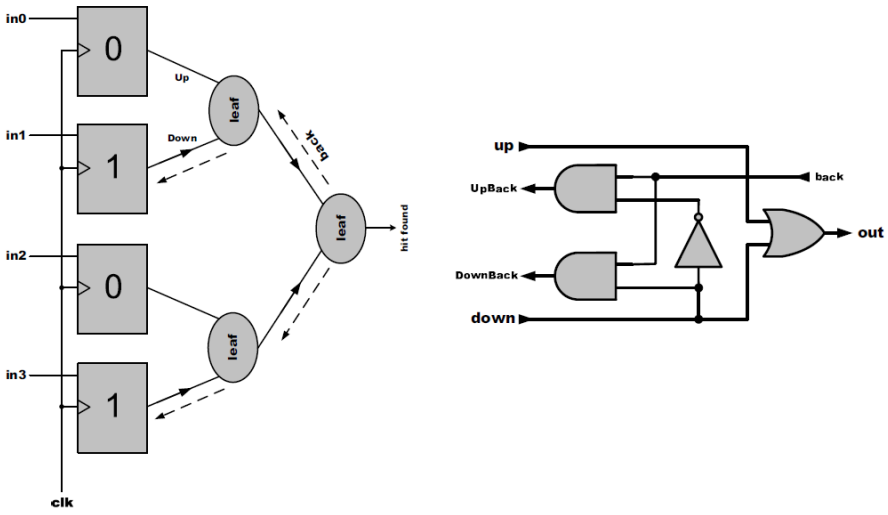


Fig. 3.7: Hitscanner design with a parallel tree structure. The existence of one or more hits is indicated by the hitflag (forward direction). One back propagting part (leaf) is shown on the right.

Testability For stand-alone testability of the analog part several test current sources for pedestal and signal are implemented in CURO 2, a global one and local ones in each channel (fig. 3.8). Each current source is adjustable by an 8-bit DAC. While the pedestals

are not switched, the signals are only active on the high level of the *WrCLK*. The active channels are selected by the so-called *hit-pattern*. For the calibration the current sources are multiplexed to the monitoring bus (*mon_out*), which is connected to an external Source Monitoring Unit (SMU).

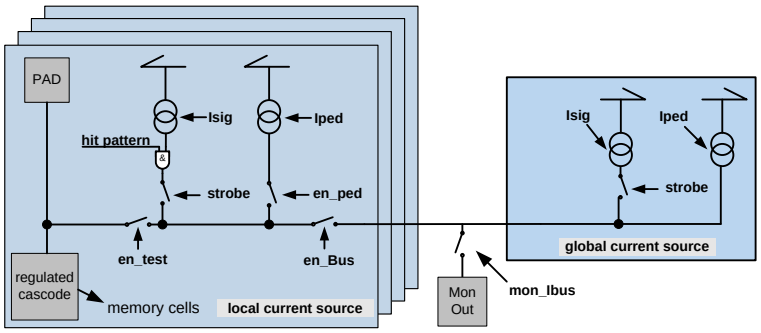


Fig. 3.8: CURO test current sources. A global test current source consisting of static pedestal and switched signal can be enabled for each channel. In addition, each channel provides a local current source.

A further feature for testability is the so-called *scan pattern*, which is a programmable pattern used as an input of the hit-scanner instead of the comparator results. Apart from functionality tests of the digital part this is used to override the comparator results in a non-zero-suppressed readout scheme.

3.5 The Hybrid

A close-up view of the chip assembly on the hybrid is shown in fig. 3.9. At the center a 64×128 pixel matrix is glued to the PCB. Two SWITCHER 2 steering chips are placed at either side, connected row-wise to the gate and clear contacts of the matrix respectively. The CURO 2 readout chip is attached at the bottom side and connected to the drains of the matrix. All electrical connections between the chips and to the PCB are made by wire bonds. Low-noise transimpedance amplifiers (AD 8015, not visible on the picture) have single-ended current input and a differential voltage output. In order to minimize stray capacitances they are used as dies (without package) and are directly wire bonded to the analog outputs of CURO.

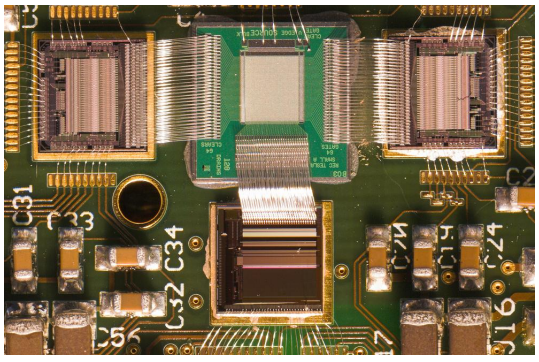


Figure 3.9: Photograph of the chip assembly on the hybrid. The sensor matrix is located in the center, steering chips are placed on either side and the readout chip is placed at the bottom.

For the wire-bond connection to the backplane the PCB has a cut-out underneath the sensitive area of the matrix. In addition, the cut-out permits laser tests on the matrix backside and minimizes the scattering material in a beam test. For flexibility at the different system setups (laser, radioactive source and beam test setup) hybrid and readout board are connected by a ribbon cable, which transmits all digital I/O signals, differential analog voltages and the supply voltages. A second, optional connector (Sub-D) is used as an independent supply to separate supply voltages and I/O signals to potentially minimize crosstalk.

3.6 DAQ board

The DAQ board (see figures 3.1 and 3.2) houses a Spartan 3 FPGA[1] (XILINX) as central component, which coordinates the synchronization and the signal processing of all components. It features 208 I/O pins, of which 173 are arbitrarily configurable. The communication with the DAQ PC is managed by a USB micro controller.
An FPGA is a digital chip containing programmable logic blocks and programmable interconnects. It is easily and repeatedly programmable. The functionality is described by a hardware description language (HDL) design, which is translated into a technology-mapped netlist by a dedicated software package and later on fitted to the actual FPGA architecture using a process called place-and-route. After validation with simulation tools, a binary file is generated, which is up-loaded to the FPGA to finish the configuration.

The design project for the prototype system includes the logic for the initialization and configuration of the ASICs, a sequencer, clock management, SRAM control, sampling of the pipelined ADCs, state machines for an automatic processing of the CURO hit data, the control of trigger and busy signals and additional testability, and debugging functionality. All different readout schemes (for zero-suppressed readout or full analog readout, for the lab environment, for beam tests, for laser, etc.) are included in a single configuration file.
The main system clock is provided by a phase-locked loop (PLL) clock, which can be adjusted up to 200 MHz in steps of 1 MHz. Due to internal timing constraints of the FPGA design, the actual system frequency is limited to 50 MHz, while the SRAM communication is done at 100 MHz.
The differential voltage signals from the hybrid are digitized by two differential 14-bit ADCs. 256 kBytes SRAM is used to buffer the processed data prior to the transfer to the PC, which corresponds to 16 full frames of non-zero-suppressed data. Levelshifters are needed to adapt the maximum possible 3.3 V signals provided by the FPGA to the required 5 V for the digital Switcher I/O. Furthermore, the board offers three freely configurable LEMO input and output ports respectively, eight arbitrary testpins to connect a logic analyzer and an I^2C interface, which can be used to connect a temperature sensor. The data transfer to the PC is done via a high speed USB 2.0 link.

[1] Field-Programmable Gate Array

3.7 Data Acquisition

For beam tests and measurements with a laser a triggered readout is the most straight forward approach. Although for the final ILC detector a triggerless readout is foreseen, in a prototype state a trigger has several advantages or is even inevitable. Due to the vast data rate a triggerless readout is only feasible with on-chip zero-suppressed readout. However, zero-suppression is not always desired, since some information, potentially important for a comprehensive system characterization, is irrecoverably lost. Furthermore, the current version of CURO lacks an analog FIFO and has a small digital buffer for only 512 hits, which reduces the zero-suppression functionality to a proof of principle.

A triggered, quasi-continuous readout is implemented with so-called *rolling shutter* mode. As long as the system is ready and waiting for triggers, the matrix rows are continuously selected and cleared one after each other. This state is referred to as *idle* state. After the last row of the array the sequence continues seamlessly with the first one, and so forth. On each trigger a frame is read out starting from the next active row on, in the following referred to as the *startgate*. Ideally, from the matrix point of view, there is no difference between the idle and the readout state. However, by choosing a higher row-rate during the idle state, the occurrence of multiple hits and the contribution of leakage current can be significantly reduced.

Since the loaded sequence runs stand-alone within the FPGA, and particularly independent from the software, the sequence is not affected by the data transfer to the PC, during which the system is in the idle state. Furthermore, the DEPFET matrix is operated under steady conditions throughout the run, as no start and stop is needed. This is unfortunately not true for the readout chip CURO, which is only activated after a trigger. The concept of the rolling shutter has been adopted to both test beam and lab measurements, independently of using the zero-suppression.

The basic process flow for a readout using the rolling shutter mode is illustrated in fig. 3.10. The scheme is in first order the same for a zero-suppressed and a non-zero-suppressed readout. In the following the zero-suppressed case is discussed, differences to the non-zero-suppressed readout will be treated later in this chapter.

In the default idle-state, the matrix is continuously cleared row by row and the readout chip is inactive (powered, but not clocked). As soon as a trigger occurs, the readout chip is activated. From the following row on, the signal currents are sampled twice with a clear in-between. Then, the pedestal subtracted analog values of that row are serially shifted out of the two CURO outputs, with each *ScCLK* two hits are processed. The currents are converted into voltages by the external transimpedance amplifiers and are digitized by the ADCs. The binary signals are written to the on-board SRAM and are transferred to the PC later on (usually when the RAM is full). As the data is buffered on the readout board, the row rate is not limited by the transfer rate. If all hits of a row are processed, i.e. no more hits are found, the next row is selected. After the frame is finished, the readout chip is again inactive, while the matrix is continuously cleared row-wise.

Triggers are only accepted if the system is not busy, which is indicated by the respective status flag. For example, triggers are not accepted during the readout of a frame, if the buffer is full or data are being transferred to the PC, but also in a test beam environment, when other devices like the telescope are busy. Consequently, the readout cycle of a frame

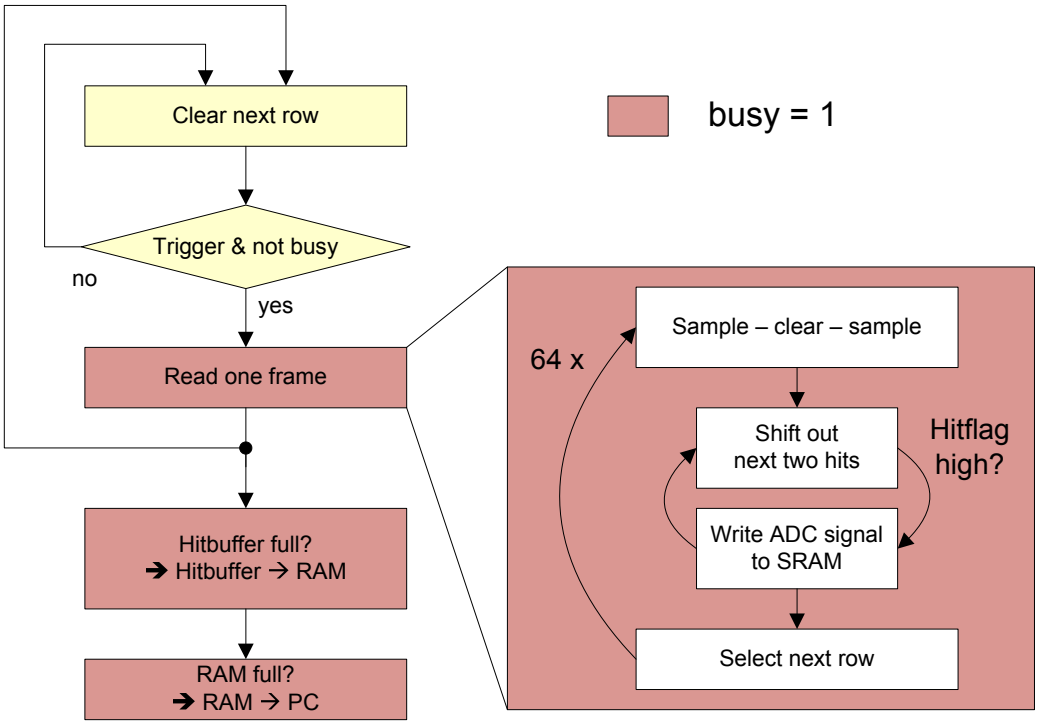


Fig. 3.10: Rolling shutter readout sequence using zero-suppression (flowchart).

is not extended in case of further raw triggers during the readout.
The process flow is mainly managed by a sequencer, which is implemented in the FPGA. It has 16 configurable output tracks, which are used for external I/O signals (for instance Switcher and CURO clocks, laser trigger and status flags) as well as for internal signals like FIFO access and storage of the ADC values to the RAM. The implementation of (embedded) loops, wait states and conditional jumps allow arbitrary readout sequences. The parameters of the sequencer are set on-line via a user-friendly software interface.

In the next section, the zero-suppressed and the non-zero-suppressed readout schemes are discussed in more detail. It will be shown later, that every processing step has an individual optimal timing, that requires to operate the CURO with single *clock strobes* rather than continuous clocks.

3.7.1 Readout modes

On-chip zero-suppressed readout For a better understanding of the zero-suppressed readout a deeper look inside the CURO is necessary. After the sampling (and pedestal subtraction) of a row, the analog signals are channel-wise compared with threshold currents (see fig. 3.11). The binary result is then being analyzed by the hit-scanner, which is steered by the Scan Clock (*ScCLK*). With the first Scan Clock cycle the hitscanner is loaded and up to two hits are identified instantaneously. The presence of at least one hit is indicated by the hit-flag. While the pixel addresses (row, column) are stored in the on-chip hit-buffer, which can hold up to 256 entries with one or two hits, the analog signals are linked to the two analog outputs of the chip instantaneously and are digitized and stored externally as usual. With each further *ScCLK* up to two further hits are processed. If no further hit is found, the hitflag goes down, the *ScCLK* is stopped and the next row is processed.
Since CURO is only activated for one frame after each trigger, the internal row index does not match the actual matrix row. Therefore, only the column information is used later. The row index is taken from a synchronized counter inside the FPGA, which reading is stored in a FIFO (implemented in the FPGA) with every hit. The CURO signal *buffer_ff* indicates if the hit-buffer is full. In this case, the addresses of further hits found by the hit-scanner are skipped. Therefore, no further triggers are accepted by the system (busy = 1) and the content of the hit-buffer is transferred to the SRAM.
To allow an automatic and continuous data acquisition, the readout is managed by a state machine implemented in the FPGA, which is triggered within the readout sequence. All reset and configuration operations and the serial data transfer of the 256×27 bits are done with full PLL frequency of up to 50 MHz, which roughly takes 140 μs. The hit-buffer data is processed before it is stored to the SRAM. The row index is read from the FIFO, the column index from CURO. To link the hits to the correct event, a frame index is stored with each hit. The re-assignment of analog and digital data is done hit by hit in chronological order with the readout of the SRAM. Care has to be taken, that one (and only one) analog value is sampled for each digitally stored hit and vice versa.
Note, that this version of CURO was not designed for a continuous zero-suppressed readout (missing analog FIFO, small hit-buffer). However, the functionality can be verified.

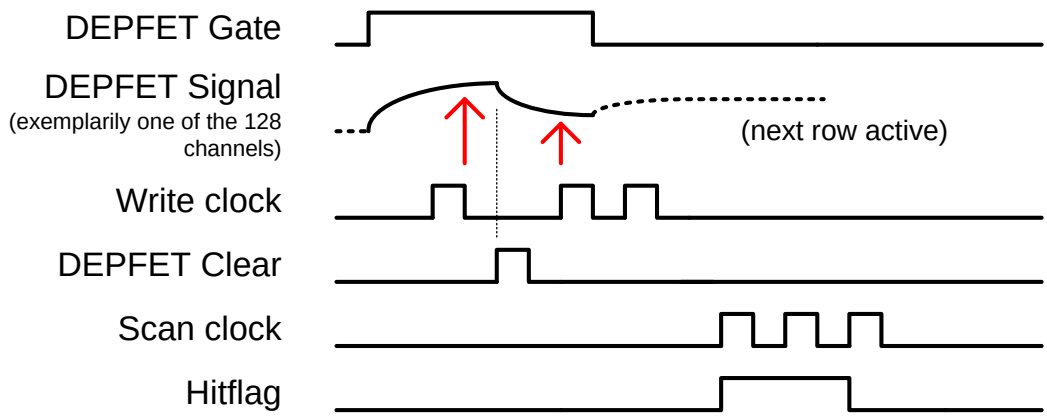


Fig. 3.11: Zero-suppressed processing scheme of a DEPFET row. After activating the DEPFET gate, the superimposed signal and pedestal current is sampled on the Write Clock (*WrCLK*) high signal and stored on the falling edge. After clearing the pixels, the pedestal current is sampled on the *WrCLK* low signal and is automatically subtracted. The 2nd and 3rd *WrCLK* strobes are needed to store the result in the analog FIFO and for the current comparator. Then, the hitscanner is loaded with the first Scan Clock (*ScCLK*) and the first two hits are identified instantaneously, indicated by the hitflag. With each further *ScCLK* cycle up to two more hits are identified. When no more hit is found, the hitflag goes down, the *ScCLK* is stopped and the next row is processed. In this example the hitflag was high for two *ScCLK* cycles, indicating that three or four hits were found. In a non-zero-suppressed readout scheme always 64 *ScCLK* strobes are given.

stage	typical time [μs]
sampling of one row (sample-clear-sample)	0.32
required row rate at the ILC	0.05
shifting out up to 2 analog hits	0.2
readout of one row	14.8
readout of one frame (non-zero-suppressed)	947.2
readout of one frame with 20 hits (zero-suppressed)	130.0

Tab. 3.2: Typical timing of the DEPFET readout with the present system. A faster readout is not possible without affecting the signal to noise ratio. A detailed analysis is discussed in chapter 4.4.

The unusual way the readout is implemented, with a buffering of the digital information only, is done in respect to the later ILC timing structure. Buffering the information of a complete bunch train for a transfer within the long bunch pause is not feasible for the analog signals. The expected number of hits read out by one readout chip (128×1000 pixels) during a bunch train in the innermost layer is roughly 34000, if an average cluster of 3 pixels per track is taken into account. Both, power consumption and the mere required space to implement such a number of current memory cells, would exceed the ILC requirements. Instead, the zero-suppressed values could either be digitized on-chip and buffered digitally, or shifted out directly without buffering. The digital information of frame number (5 bits), row (10 bits) and column (7 bits) index needs 22 bits. Hence, a buffer size of 100 kB would be sufficient for the ILC chip.

Typical durations of the individual readout steps needed with the present chip are listed in table 3.2. A detailed timing analysis is discussed in chapter 4.4.

Non-zero-suppressed readout Without using the functionality of the zero-suppression, the readout is much simplified. The results of the comparators are not used, instead the hit-scanner is loaded with a test pattern showing a hit in every channel. In this readout mode, the order of the read out pixels within a row is always the same. Thus, the association of the analog signal data to the correct pixel of the matrix can be done off-line, if the row index of the first read out row is known. Since the data in the hit-buffer is not needed, the readout can be skipped to save time. While the time needed for the sampling of a row is still the same, the serial read out of all the analog signals via only two analog outputs takes a long time (see table 3.2). The advantage of a non-zero-suppressed readout is that all information is accessible, allowing a comprehensive characterization of the system including off-line corrections of pixel offset and common-mode and different methods of cluster analysis.

3.8 Data Acquisition software framework

Beam test and lab measurements require a substantial software framework that can handle all system tests, monitoring and a continuous data taking in different environments. The software is written in C++, allowing an object oriented structure. An overview of the different applications is given in figure 3.12.
The data provided by several independent devices (particularly needed for a beam test) is combined to events by a separate application, the *Writer*. The result is a single data file containing a consistent set of events from all devices. While a stand-alone operation could also be handled by a single application, in a beam test environment a flexible event-building is necessary, where the type and amount of active devices is easily adjustable. Furthermore, it has to be assured already during the data acquisition that all devices have recorded identical physics events. If one or more of the devices processes more or less triggers than the others, the data file is corrupted.
During data taking the DEPFET software takes care of the regular readout of the acquired data from the on-board buffer. After the data processing, which includes the reallocation of the analog data to the correct pixel and the assignment of event number and additional information, the events are written to the shared buffer. These are the tasks of the *Producer*. In case of a multi-module operation, several independent instances of the producer software are executed, each labeled with its unique ID. Two top-level applications, the *Module Manager* and the *Run Control*, manage the synchronized, automated data acquisition. The communication between the different applications is done via Windows messages. An optional *Online Monitoring* can be used to check the data during the running acquisition.

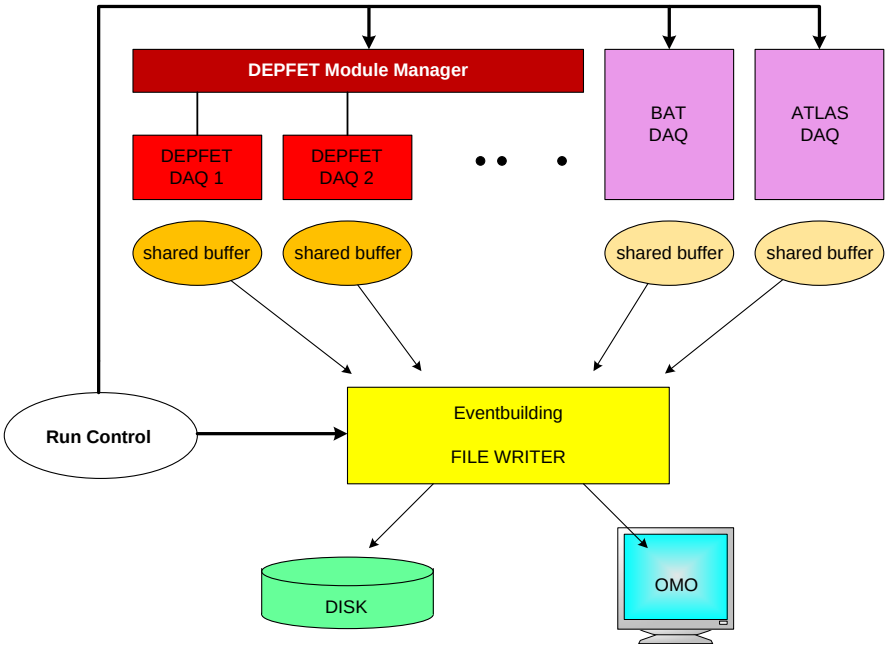


Fig. 3.12: Data acquisition software framework.

4. System characterization

In this chapter, the performance of the prototype system is presented. After a calibration of the readout chain, important parameters of the readout chip are characterized. The DEPFET sensor is characterized in terms of internal gain, leakage current and homogeneity. A detailed timing analysis of all individual readout steps and two-dimensional scans of the supply voltages are carried out in order to find the optimal operation parameters. The homogeneity of the CURO thresholds before and after tuning is analyzed with regard to the zero-suppressed readout. Furthermore, a detailed discussion of the total system noise figure is presented.

4.1 Calibration

4.1.1 Readout chain

The transimpedance amplifiers (TIAs) and the ADCs were calibrated by using the internal test current sources of the CURO. After calibrating the internal test current sources by probing the currents at the mon_out pad with an external source monitoring unit (SMU), the mon_out pad is directly connected to the TIA input by a wire bond. This is done for both analog channels separately. As shown in fig. 4.1, the following conversion factors have been observed for the tested hybrid:

$$ADC1 : (7.62 \pm 0.17) nA/ADU \tag{4.1}$$
$$ADC2 : (7.70 \pm 0.14) nA/ADU \tag{4.2}$$

The bandwidth of the ADCs is limited to 30 MHz by discrete low-pass filters. The supply voltage range of the TIA has been lowered from $\Delta U = 5.0\ V$ to $\Delta U = 3.3\ V$ to achieve an optimal S/N at the given readout speed. The input potential was adapted to the CURO output, which is approximately at $1.2\ V$, by an TIA offset of $-0.5\ V$ in respect to the system ground. As expected, the observed values are within the specifications of the components [6][7].

4.1.2 CURO

The input currents are not amplified within CURO, i.e. after pedestal subtraction and hit processing the output currents are ideally unchanged. The transfer gain has been measured for all channels by probing the output current as a function of the input current, for which again the internal test current sources are used (using a typical sample timing as in a real sequence).

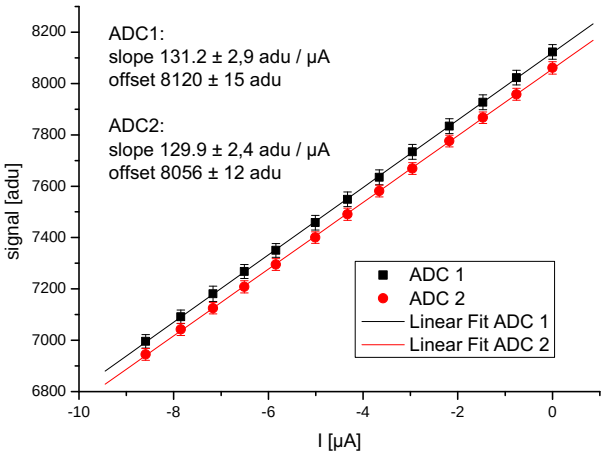


Fig. 4.1: Calibration of TIAs and ADCs using the CURO test current sources.

The transfer gain and linearity of one channel are exemplarily shown in fig. 4.2. The integral non-linearity is defined as the maximum deviation from linearity within a region of interest:

$$INL = \frac{max|y_i - y_{i,f}|}{\Delta y} \tag{4.3}$$

where y_i are the measurement points and $y_{i,f}$ numbers given by the linear fit. Assuming an internal amplification of the DEPFET sensor of 0.5 nA/e^- a dynamic range of 12 μA complies to a signal of 24000 e^-, which is a typical seed signal[1] of a MIP in the present DEPFET sensor, or 6 MIPs in a sensor of 50 μm thickness. For the required dynamic range the average integral non-linearity per channel is measured to be 0.24 %. Fig. 4.3 shows the transfer gain and the offset distribution of all channels of hybrid 2B. The average transfer gain for all 128 channels is found to be 1.036 ± 0.001, with a range of 0.2 %. The offset inhomogeneities have to be compensated by threshold tuning (see sec. 4.5). In case of a non-zero-suppressed readout they can be corrected off-line.

4.1.3 DEPFET matrix

The internal gain and the linearity of the DEPFET matrix is determined by taking the energy spectra of different x-ray sources. The dominant x-ray energies of the used radioactive sources are listed in table 4.1. The spectrum of ^{241}Am is shown in fig. 4.4 as an example, demonstrating the spectral performance of the system. The data is corrected for pedestals and row-wise common mode noise. The pixel with the highest signal above a threshold of 6 $\sigma_{pixelnoise}$ and surrounding pixels with a signal larger than $3 \cdot \sigma_{pixelnoise}$ are combined to a cluster (see chapter 5 for details on the data analysis).
In fig. 4.5 the mean signal, obtained by a Gaussian fit to the spectrum peak, is plotted as

[1] The *seed* signal is the highest signal of a cluster. See chapter 5.3 for details.

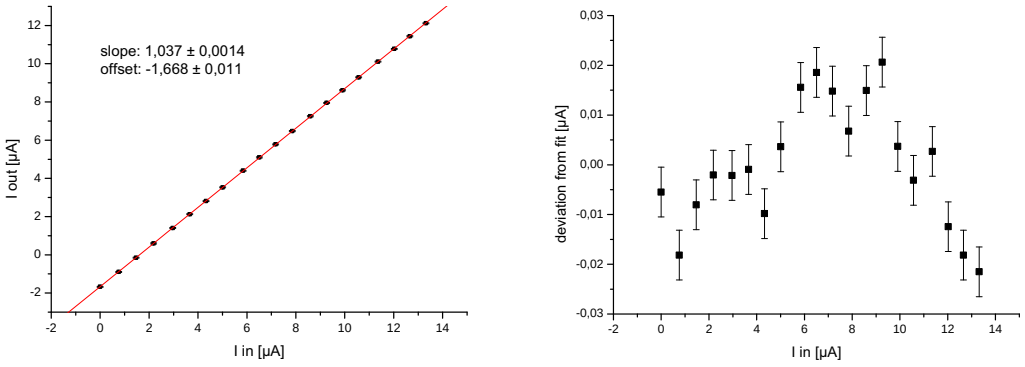


Fig. 4.2: Transfer gain and INL of channel 24, Hyb 2B (exemplary).

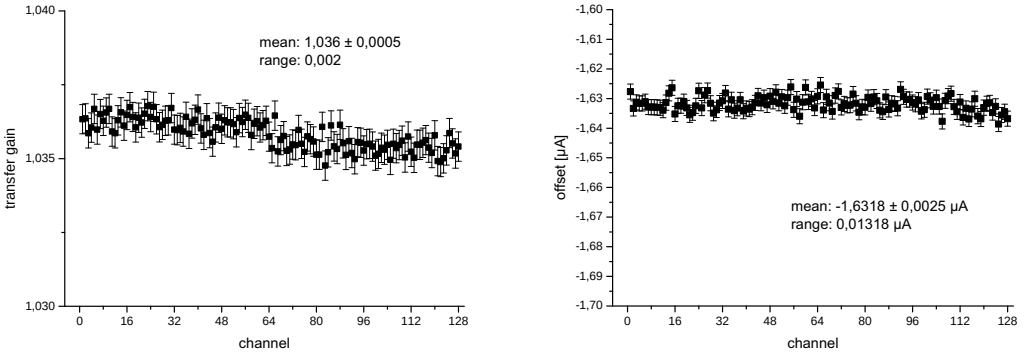


Fig. 4.3: Transfer gain and offset of the analog part for all channels of Hyb 2B. The range is defined by $y_{max} - y_{min}$.

a function of the x-ray energy of the different radioactive sources. A linear fit yields an internal amplification of the DEPFET sensor T06 (Wafer 11, high-E, common clear-gate, pixel pitch $36 \times 22\ \mu m^2$) of

$$g_Q = (308.6 \pm 6.5)\frac{pA}{e^-} \tag{4.4}$$

Note, that this kind of measurement is not sensitive to charge loss within the pixel. If a constant fraction of the signal electrons would for example drift to the clear contact, less signal charge would be accumulated in the internal gate. Since the calibration above assumes complete charge collection, the real internal gain in case of a charge loss would be higher than the observed one. The result above (4.4 is in agreement with three dimensional device simulations. However, there are strong indications for a charge loss in structures without high-E implantation, which is expected to occur mainly in the clear regions. Signal electrons generated in the fully depleted bulk are affected, if they interfere with these regions on their way to the internal gate. This effect is significantly reduced in structures with high-E implantation, since the internal gate lies more deeply in the bulk.

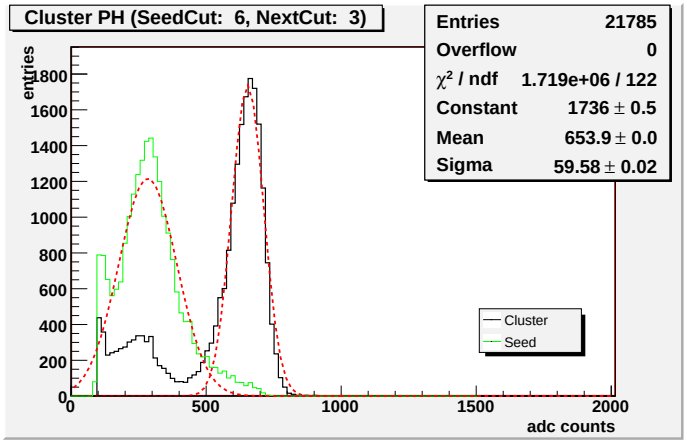


Fig. 4.4: Energy spectrum from a ^{241}Am-source. A seed cut of $6\sigma_{noisepeak}$ corresponds to roughly 100 ADU. The 59.5 keV peak is registered at 650 ADU.

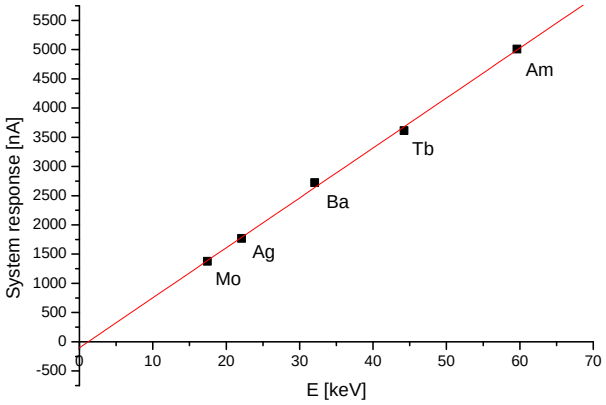


Fig. 4.5: System response for different radioactive x-ray sources.

This important issue is currently being studied with precise laser measurements and beam test analysis with high statistics.

4.2 Leakage current

During the integration time of the system, leakage current is accumulated in the internal gate and generates an offset to the signal. The statistical variation of the charge carriers causes an additional noise contribution (shot noise).
The longer the integration time between two readout processes, the more charge is accumulated in the internal gate. This signal offset is measured as a function of the integration time. The result is shown in fig. 4.6. The data is corrected for baseline pedestals, which are measured by a fast, completely pedestal subtracted sequence (clear-sample-sample).
Since leakage current shows an exponential dependence on the temperature, the temperature has been monitored during the measurement. The temperature sensor was attached

Source	Abbreviation	E_γ [keV]
Molybdenum	Mo	17.44
Silver	Ag	22.1
Barium	Ba	32.06
Terbium	Tb	44.23
Americium	^{241}Am	59.5

Tab. 4.1: Radioactive sources and the dominant emitted x-ray energies.

to the hybrid PCB, as close as possible to the DEPFET matrix. During the measurement a stable temperature of $T = 39°C$ has been observed.

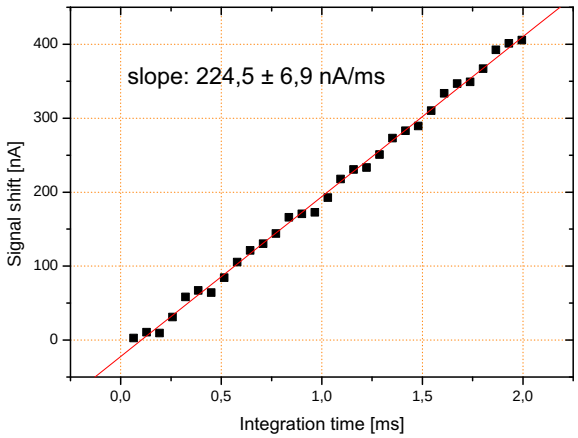


Fig. 4.6: Leakage current per pixel. The measurement took place in a light-proof box. The temperature of the matrix was $39°C$.

A linear fit to the data yields a leakage current of $I_{leak} = (116.5 \pm 5.8)$ fA/pixel for an internal gain of $g_Q = 308\ pA/e^-$, which corresponds to $I_{leak} = (14.7 \pm 0.7)\ nA/cm^2$. Given the relatively high temperature, this value is compatible with wafer scale measurements on diode test structures [32].
There are several mechanisms leading to leakage current. One is the movement of minority carriers in a reverse biased diode. Another one is the thermal generation of electron-hole pairs originating from recombination and trapping centers in the depletion region, which can catalyze the creation of electrons and holes from the valence band by serving as intermediate states. The third and by far the largest contribution is originating from surface channels, particularly in regions which are not highly doped as beneath MOS gates. With the next generation of DEPFET matrices the undoped areas are being further minimized.
The contribution of the leakage current to the signal for a typical integration time of up to 0.8 ms calculates to $\sim 600\ e^-$ (accumulated in the internal gate). This is small

compared to a MIP signal (36 ke^-) and is correctable by off-line pedestal and common mode correction. In the case of a zero-suppressed readout signal offsets would effectively result in a threshold shift (common signal higher, thresholds effectively lower). However, due to the much higher row rate the accumulated leakage charge is negligible.

4.3 Homogeneity

The inhomogeneities of the DEPFET pixel pedestals are compensated by the on-chip pedestal subtraction of the readout chip. The chip must therefore be able to cope with the given pedestal range of the DEPFET array. Ideally, the output of the analog stage is independent of the pedestal value for a given signal. Inaccuracies of the pedestal subtraction contribute to the noise. In case of a non-ideal pedestal subtraction a higher threshold is required for hit discrimination to compensate pedestal variations.
The homogeneity of the DEPFET pixel array is difficult to measure directly with the existing setup, since the on-chip pedestal subtraction cannot be deactivated.
By studying the accuracy of the pedestal subtraction and the remaining pedestals of a DEPFET matrix after subtraction, an estimation of the homogeneity can be given. The performance of the pedestal subtraction has been measured by sweeping the on-chip pedestal test current sources for a constant signal. As reported in [39], a deviation from linearity of (1.53 ± 0.01) % has been observed. The remaining pedestal distribution of a DEPFET matrix is shown in fig. 4.7. The performance is highly dependent on the timing (see next section). Using a reasonable timing yields an RMS over the whole matrix of (55.6 ± 0.1) nA, demonstrating that the homogeneity is within the expected range. Note, that this number implies also all contributions of the readout chain behind the pedestal subtraction stage. The value given here is thus an upper limit.

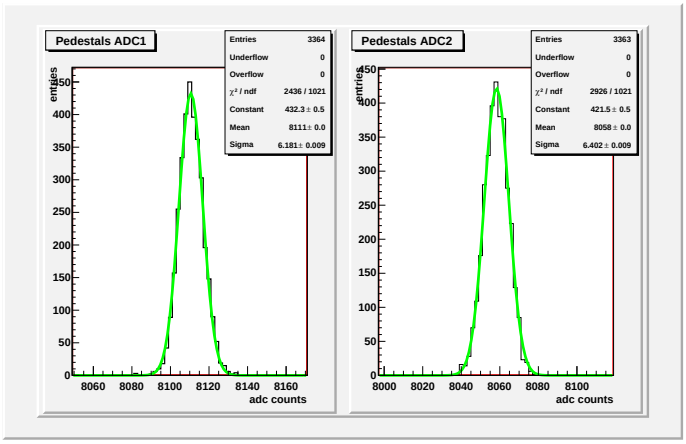


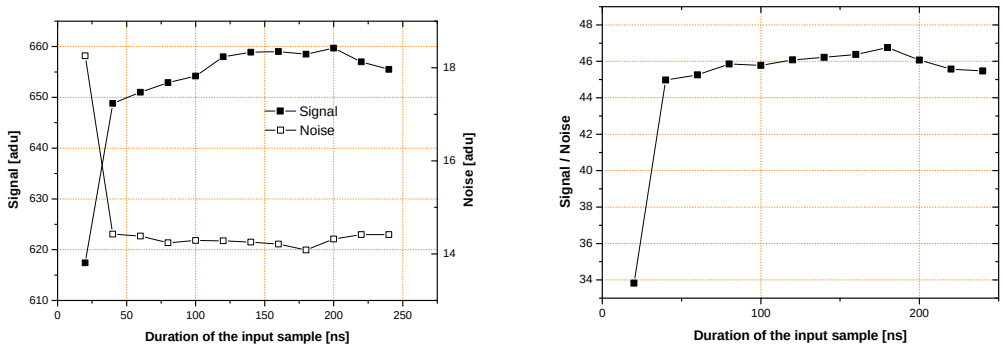
Figure 4.7: Homogeneity of the DEPFET pedestals after on-chip pedestal subtraction.

The gain homogeneity of the different matrix designs is currently under study [31]. Preliminary measurements show a satisfying homogeneity with deviations smaller than 5 %.

4.4 Readout timing

The system clock is provided by an adjustable PLL with a maximum frequency of 50 MHz. This clock is used for the entire readout, which is managed by the sequencer. The timing of the individual components like SWITCHER and CURO can thus be adjusted in steps of 20 ns.
The data storage of both 14-bit ADCs to the SRAM needs to be done within one clock cycle. By processing the ADCs one after each other the amount of required I/O ports is reduced to the half. In turn, the clock rate must be twice as fast as the system clock rate, i.e. a maximum of 100 MHz.
Since the signal data is buffered on-board, the transfer rate to the PC does not limit the line rate of the readout. The continuous data taking has to be interrupted if the SRAM is full. This happens after 16 complete frames. The 256 kBytes of data are in this case transferred to the PC, which takes approximately 20 ms via the USB 2.0 link.
Besides the data transfer the actual readout speed of the current prototype system is limited mainly by the readout chip. In the following, the timing of the two CURO clocks is discussed in more detail (see chapter 3.4 for a description of CURO 2 and figure 3.11 for the readout scheme). The measurements have been performed using an ^{241}Am γ-source and DEPFET hybrid 2B.

Input sample (1st *WrCLK*) During the high-phase of the *WrCLK* the signal plus the pedestal value is sampled. In fig. 4.8 the signal is shown as a function of the duration of the sample signal. A sufficient settling is reached within $40 - 60$ ns.

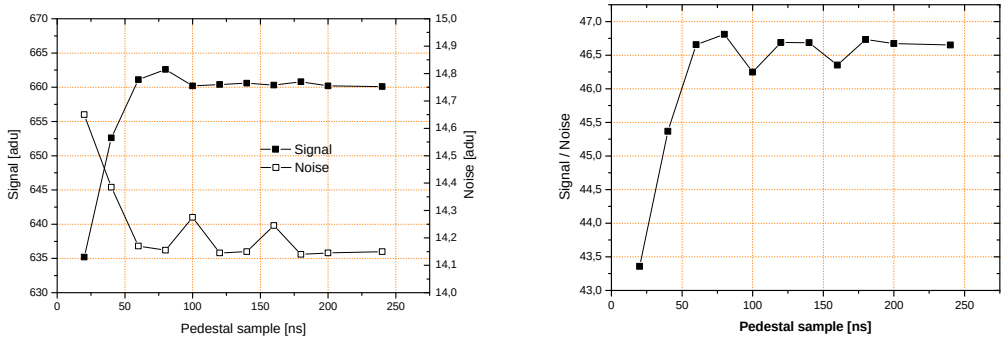


(a) Signal and noise.

(b) Signal to noise ratio.

Fig. 4.8: Signal and noise as a function of the duration of the input sample (signal + pedestal).

The low phase of the *WrCLK*, during which the pedestal after charge removal from the internal gate is sampled, shows similar behavior (see fig. 4.9).
However, higher than average pedestal values as observed in some parts of the matrix are not sufficiently subtracted, yielding a signal offset in the respective parts. In those areas a significantly longer pedestal sample of at least 100 ns is needed. The reason for this is



(a) Signal and noise.

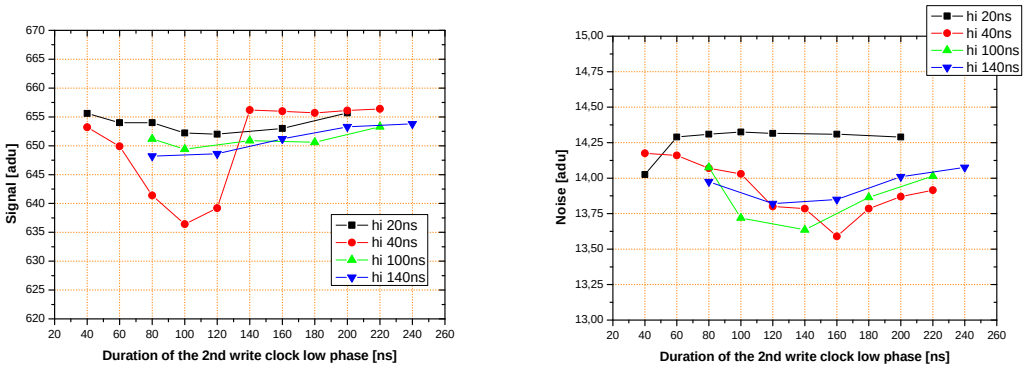
(b) Signal to noise ratio.

Fig. 4.9: Signal and noise as a function of the duration of the pedestal sampling pulse.

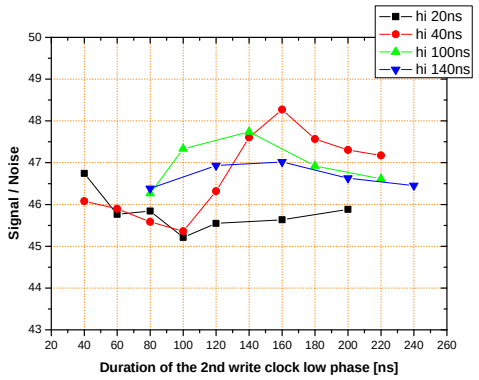
the coarse sample stage of the pedestal sampling, which is only active for a maximum of 15 *ns* after the falling edge of the *WrCLK*. However, this time is needed for a minimum clear pulse of 20 *ns*. Thus, the coarse sample is effectively not used and the entire sample process relies on the fine sampling stage, which is not only much slower than the coarse sampling stage, but has also to be corrected for the wrong coarse value.
This deficiency is not observed when using the signal source implemented in the CURO, where no clear is needed and perfect timing in respect to coarse and fine sampling of the memory cells can be achieved.

FIFO sample (2nd *WrCLK*) During the second *WrCLK* phase the signals of the sampled row are stored into the analog FIFO memory cells. Furthermore, the comparator is pre-charged to the trip-point during the low-phase, which causes a maximum power consumption of the inverter stage. The observed signal to noise ratio as a function of the clock rate shows an optimum at a clock period of 200 *ns* with a high phase of 40 *ns* (see fig. 4.10). The characteristics is influenced mostly by the noise, showing an optimum at a duration of the low phase between 120 and 160 *ns*, which seems to be relatively independent of the duration of the high phase of the *WrCLK*.
Interesting is the fact, that particularly the low phase of the clock influences the noise behavior. While the FIFO is sampled during the whole clock period, only during the low phase the comparator pre-charge is active. Since the comparator is not used with the non-zero-suppressed readout an influence of the low phase of the clock is not expected. However, the analog power consumption increases significantly (plus 50 %) with the pre-charge. That the observed behavior is related to fluctuations on the supply voltage is therefore most likely.

The influence of the second *WrCLK* on a zero-suppressed readout is stronger, since the threshold is very sensitive to fluctuations on the supply voltage due to the pre-charge mechanism. In fact, it was difficult to find stable operating conditions, which impeded a detailed timing study. It turned out, that a longer pre-charge generally yields a higher



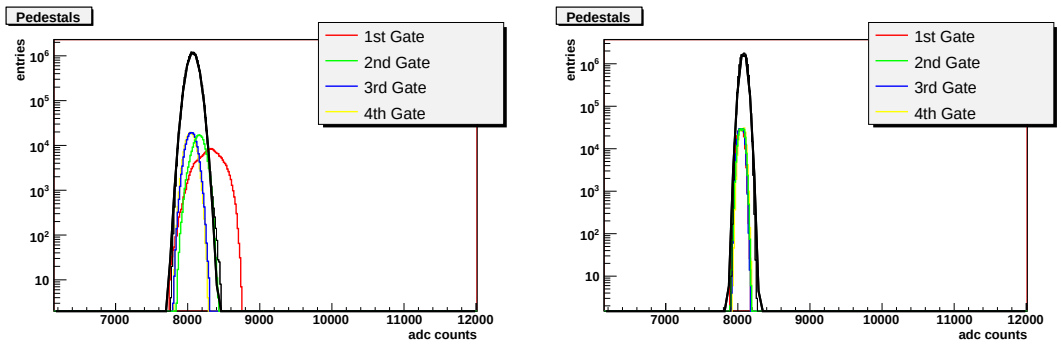
(a) Signal. (b) Noise.



(c) Signal to noise ratio.

Fig. 4.10: Signal and noise as functions of the duration of the second *WrCLK* (FIFO sample).

performance.



(a) (b)

Fig. 4.11: Raw pulse height distributions with (a) a FIFO sample of 280 ns and (b) a longer FIFO sample of 420ns.

In addition to the described behavior a significant increase of the row-wise common mode noise is observed with faster timing of the FIFO sample. A peculiarity is the so-called *first row effect*, which describes the systematic shift of the analog signal within the first few rows read out. In fig. 4.11(a) and (b) the signal spectra are plotted individually for the first four rows which are read out. As a reference the signal distribution of all other pixels is shown. The shift especially of the first row to higher values is clearly visible in fig. 4.11(a). The effect can be avoided by a significantly longer duration of the FIFO sampling (see fig. 4.11(b)), but the underlying reason for this shift is not fully understood. A reasonable assumption is that the issue is related to the readout scheme. As the readout chip is not clocked continuously, but is only activated upon a trigger a drift of the memory cells during these pauses is possible. Testing this hypothesis by continuously clocking the *WrCLK* during the idle state does, however, not solve the problem.

Comparator sample (3rd *WrCLK*) The comparator sample is performed during the high phase of the third *WrCLK*. Since no process in the analog regime is involved, the influence on signal and noise with a non-zero-suppressed readout scheme is negligible (fig. 4.12), as expected. The impact on the performance of the hit discrimination was studied using the internal test current sources. Also here no dependency has been observed, indicating that the response time of the inverter stage is below 20 ns.

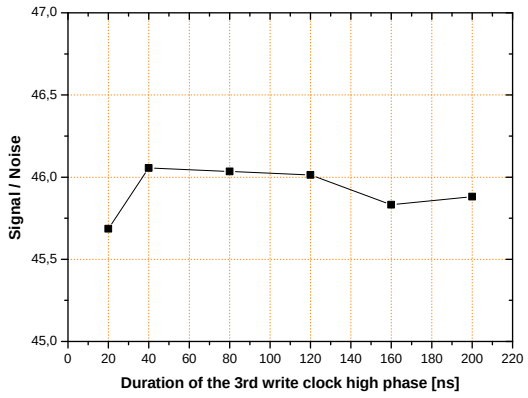


Figure 4.12: Signal to noise ratio as a function of the duration of the comparator sample.

Output sample (*ScCLK*) After processing the signals up to the hitfinder the signals of the identified hits are subsequently multiplexed to the two single ended analog outputs of CURO. The signals are valid with each rising edge of the *ScCLK*. In figure 4.13 the settling behavior of one CURO output is shown exemplarily. The oscillations after the *ScCLK* rising edge limit the readout speed considerably. This is shown in fig. 4.14, where the signal and noise are plotted as a function of the time between *ScCLK* and the ADC sample trigger. A sufficient settling is reached after $120 - 160$ ns, which is much slower than required.
The transimpedance amplifier (Analog Devices AD8015) is specified to have an analog bandwidth of 240 MHz [6] at the nominal supply voltage of 5 V, which is more than

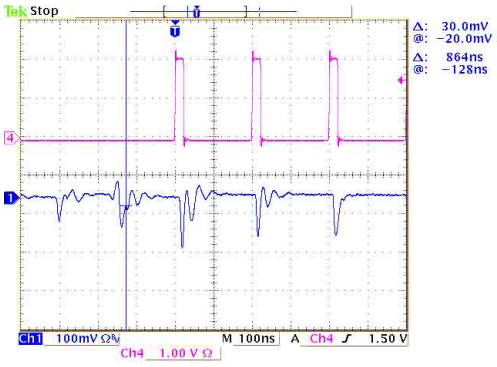
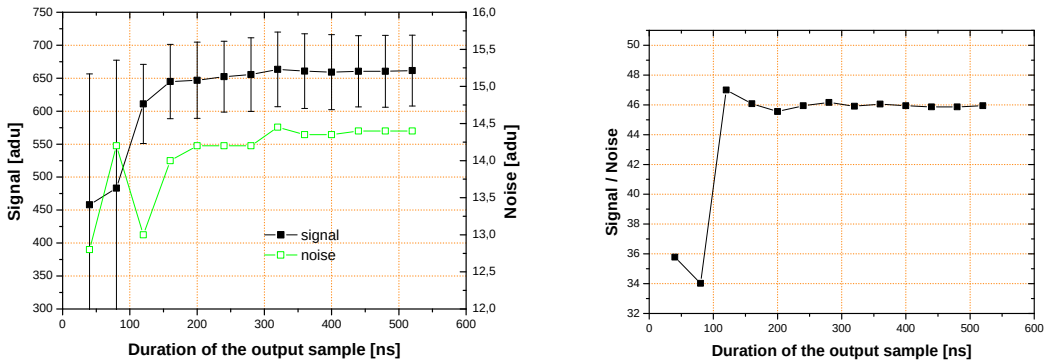


Figure 4.13: CURO analog outA (lower curve) and *ScCLK* (upper curve). The oscillations after the *ScCLK* rising edge considerably limit the readout speed.

sufficient even for the envisaged high readout rate. Although the supply voltage is reduced to 3.3 V the bandwidth should be high enough. Also the ADC (Analog Devices AD9244), which is externally limited to 30 MHz for a low noise operation, provides a sufficiently high bandwidth. Since the transimpedance amplifier is designed for very low input capacitances of photodiodes in fibre optic receiver circuits, the relatively high capacitance of the CURO output bus may be critical.



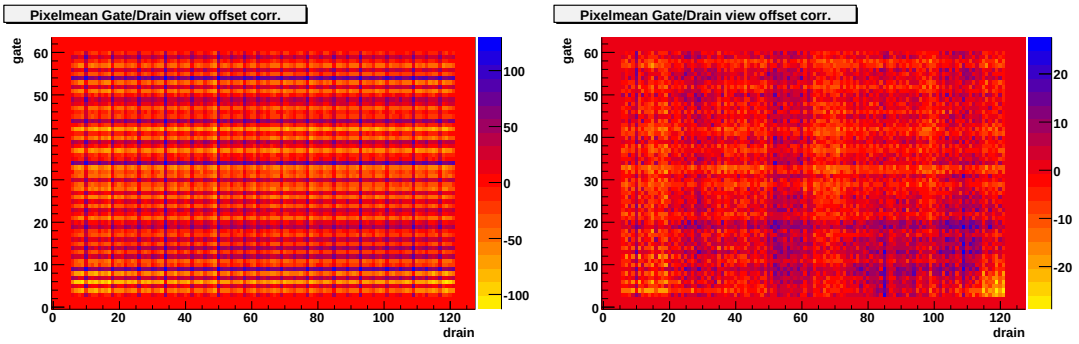
(a) (b)

Fig. 4.14: Performance of the output sample. Signal, noise (a) and signal to noise ratio (b) are shown as a function of the output sample timing.

Using a reasonable timing based on fig. 4.14 (in the order of 200 ns), a highly regular pattern of higher and lower pedestal values within the pixel array is observed (see 4.15(a)), which is stable throughout a run of data acquisition. The homogeneity is significantly improved by using an even slower timing, as shown in fig. 4.15(b) (note the different scaling). The reason for this effect is not fully understood. The regularity of the pattern, which shows higher and lower pedestal signals in every second or fourth row and column respectively, suggests a transient effect or a crosstalk of a counter to the analog output. In case of a non-zero-suppressed readout this pattern is only a minor issue, since the pedestals are off-line corrected[2]. Unfortunately, the pattern is not predictable throughout several runs. Thus, an off-line correction is not feasible in case of a zero-suppressed readout

[2] It has been shown that the superimposed signal itself is not affected

scheme, and the analog performance is directly affected. However, since the pattern is generated after the comparator stage, the hit discrimination is not degraded.



(a) (b)

Fig. 4.15: Pedestal pattern for different output sample timings. The pedestal values are shown color-coded in ADU. The homogeneity is significantly improved by a slower timing (b, note the different scaling).

4.5 CURO Thresholds

The thresholds of the 128 channels are set by a global 8-bit threshold DAC, which has a design step size of 0.7 μA. To compensate the process-related dispersion the thresholds are individually tuneable by a 5-bit tune DAC. The step size (least significant bit, LSB) of the tune DAC can be adjusted to the required range (span by the maximum deviations from the mean threshold) by an additional 8-bit trim DAC (see fig. 4.16). This layout allows a flexible adjustment with a maximum precision.
As discussed in sec. 4.6, the trim DAC threshold scan is a very sensitive tool to determine the noise contribution of the readout chip up to the comparator stage. The relevant DACs thus have to be calibrated to allow a precise conversion to currents.

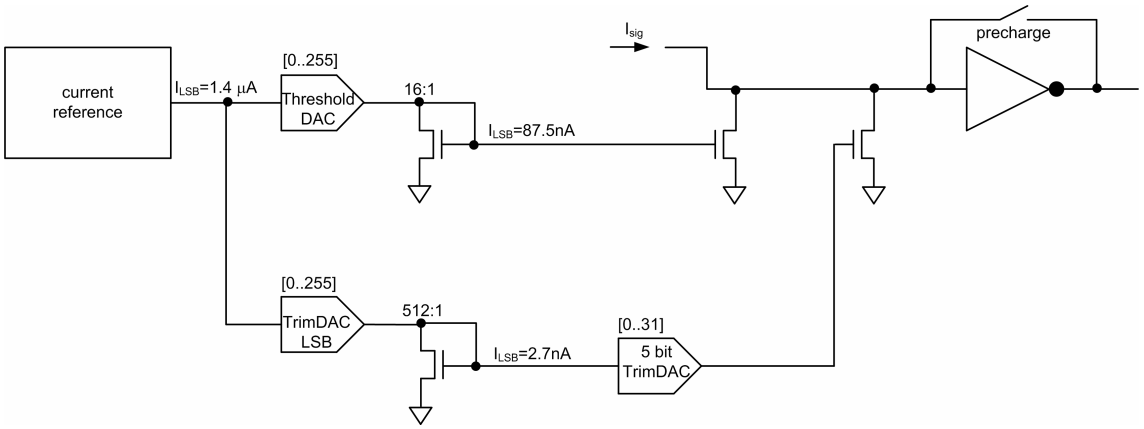


Fig. 4.16: Schematic of the comparator stage.

Calibration of the internal DACs There are several DACs implemented in CURO, which are listed in table 4.2. The DACs are realized following a common design. One global current source is used as a reference for all DACs on the chip. As shown exemplarily for the threshold DACs in fig. 4.16, the output current of each DAC is reduced by a dedicated current mirror to reach small currents and a high linearity. The design reduction ratios of the DACs are listed in table 4.2. For testing purposes the outputs of the DACs can be multiplexed to the mon_out pad individually. However, the currents behind the current mirrors cannot be measured directly, except for the signal current. Therefore, the calibrations of threshold DAC and trim LSB DAC are done indirectly using the calibrated signal source and:

$$I_{thresh} = \alpha \cdot threshDAC + \beta \cdot trimDAC_{LSB} \cdot tuneDAC \tag{4.5}$$

where α and β are the reduction ratios to be measured. As there is no current mirror after the tuneDAC, the tuneDAC is assumed to be perfectly linear. In the following the calibrations of the threshold DAC and the combination of 5-bit tune DAC and 8-bit trim LSB DAC are discussed.
According to the design one signal DAC step equals the same current as eight threshold DAC steps. The threshold shift was therefore measured for different signal currents. The result is shown in fig. 4.17. A linear fit yields a conversion value of $\alpha = (71.6 \pm 2.2)$ nA/thresh DAC step, which complies with a reduction ratio of 1 : 19.6.

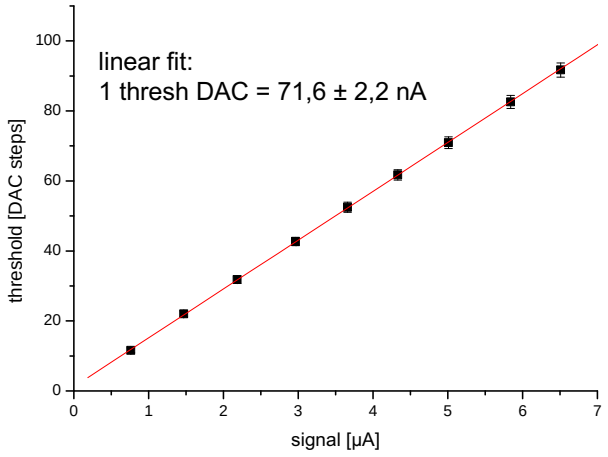
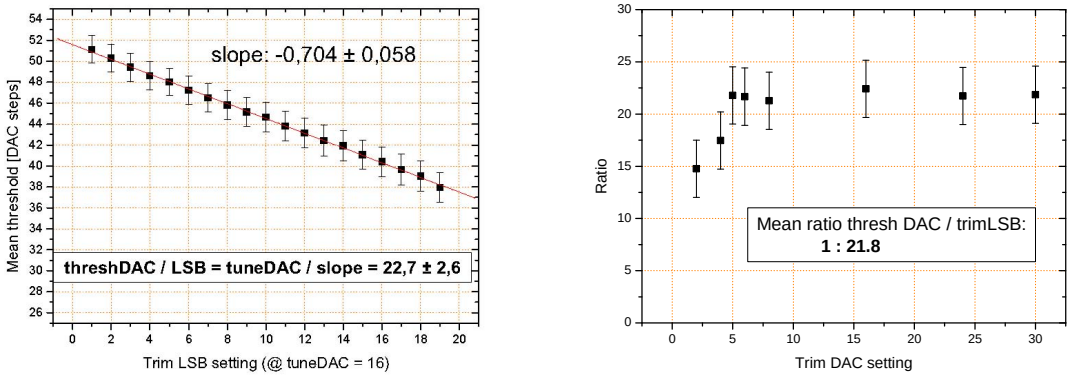


Fig. 4.17: Calibration of the threshold DAC using the internal signal source. A linear fit yields a conversion ratio of (71.6 ± 2.2) nA per DAC step.

Analogous measurements were done to calibrate the trim LSB DAC. The threshold shift has been measured for different settings of tune DAC and trim LSB DAC. An average ratio of 1 : 21.8 has been observed (fig. 4.18(b)), as shown exemplarily for a trim DAC setting of 16 in fig. 4.18(a). Taking the deviation of the thresh DAC reduction ratio from the design value into account, this complies with a conversion value of $\beta = 3.28\ nA/trimLSBDACstep$ instead of the design value of $\beta = 2.7\ nA$.



(a) (b)

Fig. 4.18: Calibration of the trim LSB for different trim DAC settings. In (a) the threshold shift as a function of the LSB of the trim DAC for a tune DAC setting of 16 is shown. The slope yields a ratio of 1 : 22.7. An average ratio of 1 : 21.8 is observed (b).

DAC number	biasing value	ratio	I_{LSB}	range
8	Iped	1:2	$0.7\mu A$	$180\mu A$
9	Isig	”	”	”
10	Iin	1:16	$87.5nA$	$22.5\mu A$
11	Ithresh	”	”	”
12	Itrim, LSB	1:512	$2.7nA$	$700nA$

Tab. 4.2: Overview of selected DACs (8 bit) and their corresponding biasing values, ratios and ranges.

Threshold tuning The homogeneity of the thresholds of the 128 channels have been measured by sweeping the threshold settings at a given signal test current or vice-versa. The response of a comparator is either 0 (or 1) for signals smaller (or larger) than the threshold. This step function is folded by a gaussian distribution, resulting from the noise of the signal source, two current memory cells and the threshold current source. The threshold is defined as the value with a probability for a positive response of the comparator of 50 %. A typical result of one channel is shown in fig. 4.19. The function

$$y = \frac{1}{2}\left[1 - erf\left(\frac{x - x_0}{\sqrt{2}\sigma}\right)\right] \tag{4.6}$$

with the error function defined as

$$erf(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-u^2} du \tag{4.7}$$

is fitted to the data.

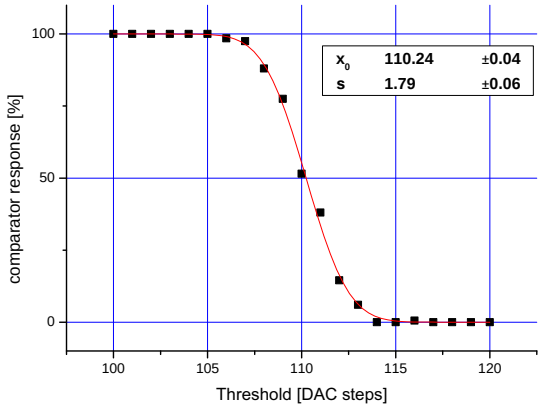


Fig. 4.19: Comparator response as a function of the threshold DAC, fitted by an error function.

The threshold dispersion of the chip before tuning is found to be $\sigma = (170.5 \pm 16)\ nA$ (as shown in figure 4.20).
The tuning algorithm implemented in the software calculates the maximum current required to compensate for the threshold dispersion:

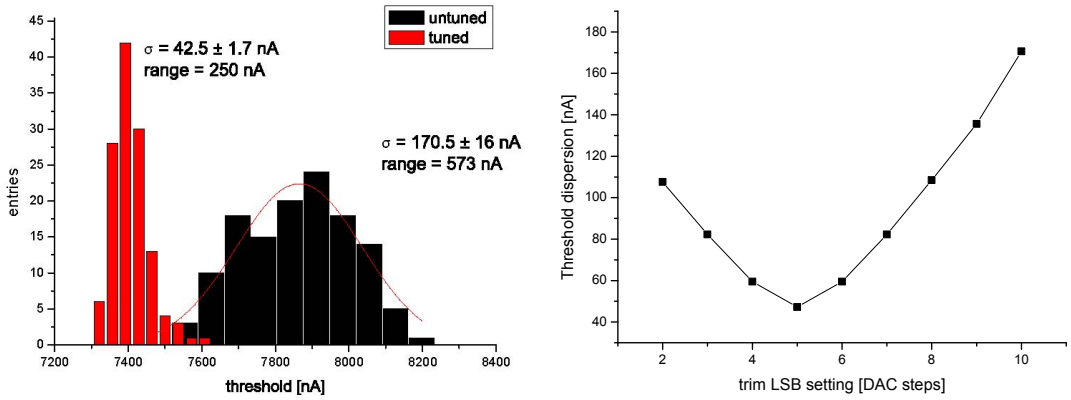
$$I_{tune,max} = I_{thresh,max} - I_{thresh,min} \tag{4.8}$$

To compensate the measured range $I_{tune,max} = 573\ nA$ a global trimDAC setting of $573nA/(31 \cdot 3.3nA) = 5.6$ is required. The thresholds are then tuned to the value of the maximum threshold by setting the channel-wise 5-bit tuneDACs. After tuning with the calibrated LSB DAC the dispersion is reduced to $\sigma = (42.5 \pm 1.7)\ nA$. As shown in figure 4.20(b) the minimum dispersion is achieved with a trimDAC setting of 5, which is in agreement with the calculated value. Given the internal gain of the DEPFET sensor of $g_Q = 0.3\ nA/e^-$, the threshold uncertainty complies to an $ENC = (138.0 \pm 5.5)\ e^-$.
It turned out that the stability of the thresholds depends on several conditions, which complicates applying these thresholds in real data taking. In particular the *WrCLK* timing is delicate, since almost all analog operations are derived from it. Furthermore, the *WrCLK* also switches the pre-charge of the comparator and the discrimination. In particular, if the clock rate is not constant but depends on the number of detected hits, as in the case of a zero-suppressed scheme described in 3.7.1, fluctuations of the thresholds are observed. This is discussed further in chapter 5.

4.6 Noise performance

In this section an estimation of the expected total system noise is presented and compared to the measured values of the beam test data. Thus, for the calculations the parameters set during the beam test are used. The following noise sources are taken into account:

- DEPFET: Shot noise due to leakage current, thermal noise and 1/f noise in the transistor channel.



(a) Threshold dispersion

(b) Threshold dispersion as a function of the trim LSB.

Fig. 4.20: Threshold dispersion before and after tuning (a). Plotted as a function of the trim LSB, the dispersion is found to be minimal at a value of 5 DAC steps as expected.

- CURO: Sampling noise (kT/C) of the memory cells.
- SWITCHER: Thermal noise of the pass transistor (kT/C nature).
- Transimpedance amplifier: Shot noise as stated in the data sheet [6].

The noise contributions are converted into the equivalent noise charge (ENC), which describes the noise in terms of the charge at the detector input (here: internal gate) needed to create the current output corresponding to the noise current:

$$\langle ENC^2 \rangle = \frac{\langle I_D^2 \rangle}{g_Q^2} \quad \text{, where} g_\mathrm{Q} = \frac{\delta \mathrm{I_D}}{\delta \mathrm{Q_{iG}}} \tag{4.9}$$

A voltage noise source at the external DEPFET gate is converted into the ENC by using the transconductance g_m and the amplification of the internal gate g_Q:

$$\langle ENC^2 \rangle = \frac{g_m^2}{g_Q^2} \langle V_G^2 \rangle \tag{4.10}$$

DEPFET 1/f noise Low-frequency 1/f noise is caused by trapping and detrapping processes particularly near the Si-SiO_2 interface of the DEPFET transistor. The 1/f noise can be suppressed by a fast correlated double sampling (see chapter 2.3.1) in the readout chip.
The ENC of low frequency noise of a DEPFET sensor is given by (from [39]):

$$\langle ENC^2_{DEPFET1/f} \rangle = a_{1/f} \frac{g_m^2}{g_Q^2} \cdot 2 \int_0^\infty \frac{1 - cos(2\pi\nu_c\tau \cdot x)}{x(1+x^2)} dx \tag{4.11}$$

The integral is solved numerically using the following values:

- $1/f$ coefficient $a_{1/f} = 1.8 \cdot 10^{-11} V^2$ (from [39]),

- system bandwidth $\nu_c = 30 \cdot \pi/2$ MHz,
- an interval for the consecutive samples $\tau = 400\ ns$.

This leads to

$$ENC_{DEPFET1/f} = (1.7 \pm 0.1)e^- \tag{4.12}$$

DEPFET shot noise The origin of shot noise are statistically independent generation and recombination processes in reverse biased diodes [36]. The spectrum of the noise source is white.
In the case of DEPFET shot noise leads to a fluctuation of charge carriers accumulated in the internal gate during the integration time.
The current spectral density is given by:

$$\langle i^2_{shot} \rangle = 2e\langle i_{leak} \rangle \mathrm{df} \tag{4.13}$$

This translates to

$$\langle \mathrm{dq}^2 \rangle = 2\mathrm{e}\langle \mathrm{i_{leak}} \rangle \Delta \mathrm{t} \tag{4.14}$$

The square root of (4.14) is the equivalent noise charge ENC:

$$ENC_{DEPFETshot} = \sqrt{2e\langle i_{leak} \rangle \Delta t} \tag{4.15}$$

With a leakage current of (117 ± 6) fA/pixel and an integration time of $\Delta t = 896\ \mu s$ this translates to

$$ENC_{DEPFETleak} = (36.1 \pm 8.5)e^- \tag{4.16}$$

DEPFET thermal noise Thermal motion of the charge carriers inside an electrical conductor lead to the so-called thermal noise. For an ohmic resistor the current spectral density $\langle i^2_{th} \rangle$ is given by

$$\langle i^2_{th} \rangle = \frac{4kT}{R} df \tag{4.17}$$

Translating this expression to a transistor channel leads to:

$$\langle i^2_{th} \rangle = \gamma g_m 4kT df \tag{4.18}$$

where γ is a semi-empirical constant that depends on the carrier concentration in the channel and the device geometry. It is usually given as 2/3 for long channel MOSFETs. With a measured g_m of the device at $V_{GS} = -4V$ and $V_{DS} = -5V$ of $(37 \pm 2)\ \mu A/V$ at room temperature and a bandwidth of 30 MHz (low pass filter) this translates to a current noise of $\sqrt{\langle i^2_{th} \rangle} = (4.4 \pm 1.0) nA$, and the corresponding equivalent noise charge due to equation (4.9) is:

$$ENC_{DEPFETthermal} = (14.2 \pm 3.3)e^- \tag{4.19}$$

CURO sampling noise The dominant noise contributions from the readout chip CURO have been simulated by transient noise analysis, where random noise is generated for every device at every time step [21]. Besides the simulation of a single memory cell (mc) three memory cells have been connected exactly as implemented in CURO readout chain (see [39] p. 117). It has been found, that thermal noise and kT/C switching noise contribute roughly equally. Flicker noise of the biasing circuits can be neglected. The total current noise has been calculated to be 92 nA, if a 30 MHz low pass filter (as implemented on the prototype hybrid in front of the ADCs) is used at the output. This translates to an

$$ENC_{3\,mc,\,calulated} = (299 \pm 60)e^- \tag{4.20}$$

Due to the difficulties to simulate annular transistors and the underlying assumptions of the simulations (input current $I_{in} = 0$, constant clock frequency of $f = 6.66$ MHz), which do not perfectly match the real operating conditions, a large uncertainty of 20% has to be assumed [21].
The noise contribution of a memory cell has been measured by probing the comparator response for different threshold settings at a constant signal current (trim DAC scan) [**?**]. The width of the resulting threshold curve (error function) results from the noise contributions up to the point of the comparator, which includes two memory cells, the signal and the threshold current sources. For different measurement series noise values between (81 ± 14) nA and (102 ± 11) nA have been observed. These values translate to an ENC between $ENC_{1\,mc} = 170.3e^-$ and $ENC_{1\,mc} = 221.9e^-$[3] for a single memory cell. For an approximation of the total noise of the readout chain comprising of three memory cells this value is multiplied by a factor of $\sqrt{3}$, leading to an $ENC_{3\,mc,measured} = 295.0e^- - 384.4e^-$. The significance of these values is limited due to a high uncertainty.

SWITCHER thermal noise The dominant noise contribution of the steering chip to the system noise figure is thermal noise of the pass transistor used to provide the switching voltages of the external DEPFET gate. The equivalent noise charge is calculated as [39]:

$$\begin{aligned} \langle ENC^2_{steer} \rangle &= 2 \cdot \frac{kT}{C_{row}} \frac{g_m^2}{g_Q^2} (1 - e^{-\tau / R_S C_{row}}) \\ &\approx 2 \cdot \frac{kT}{C_{row}} \frac{g_m^2}{g_Q^2} \qquad \text{for} \quad \tau \geq 2\pi R_S C_{row} \end{aligned} \tag{4.21}$$

with the switch resistance R_S, the capacitive load of the matrix row C_{row} and a sampling interval τ. For a typical value of $C_{row} \approx 15$ pF of the present DEPFET sensor, this translates to:

$$ENC_{steer} = 2.8e^- \tag{4.22}$$

[3] The noise contribution per current source is approximated to be 23 nA [39].

noise source	ENC (estimates)
DEPFET 1/f noise	$(1.7 \pm 0.1)e^-$
DEPFET shot noise	$(36.1 \pm 8.5)e^-$
DEPFET thermal noise	$(14.2 \pm 3.3)e^-$
CURO sampling noise	$(299 \pm 60)e^-$
SWITCHER switching noise	$\approx 2.8e^-$
TIA	$(87.0 \pm 4.3)e^-$
total noise	$\approx (314 \pm 60)e^-$

Tab. 4.3: Calculated noise contributions of the present DEPFET system.

Total noise The calculated noise contributions are listed in table 4.3. The readout chip CURO is by far dominating the total noise figure. Since the current noise contribution of the readout chip scales with the internal amplification of the DEPFET sensor, the noise figure can be improved by a higher g_Q.
The measured noise figure of $ENC_{total} = 318\ e^-$ is in agreement with the calculated value. However, the large uncertainties of the noise contribution of the readout chip weaken the strength of this comparison. Furthermore, the effects observed in the beam test data, which are most probably due to inherent fluctuations of the supply voltages are not considered in the calculations.

4.7 Optimization of operating parameters

To optimize the S/N and to ensure a complete clear, two-dimensional parameter scans were made. For the charge collection the most important parameters are the Clear-Gate and the Clear low potentials. For the clear operation Clear-Gate and Clear high are the key parameters. A complete clear in first order leads to a high signal (pedestal subtraction with empty internal gate) and low noise.
For most of the measurements presented in this thesis Hybrid 2B was used, hosting a matrix with common clear-gate (CCG), high-E implantation and pixel sizes of $36 \times 22\ \mu m^2$. Since a smaller matrix with the same geometry suited for the dedicated measurement setup with discrete readout channels was unfortunately not available, the clear efficiency could not be verified experimentally. However, there are strong indications for a complete clear resulting from studies with similar layout parameters.
These studies lead to a set of default operating parameters for each type of matrix design. The results for hybrid 2B are listed in table 4.4.

4.8 Summary

In this chapter the basic characteristics of the DEPFET ILC prototype system have been presented. The properties of the first ILC type DEPFET sensor (PXD-4) have been

Contact	Potential [V]	Contact	Potential [V]
Source	7.0	$Gate_{ON}$	-4.0
Bulk	10.0	$Gate_{OFF}$	6.0
Backplane	-180.0	$Clear_{HIGH}$	15.0
		$Clear_{LOW}$	2.0
		Common Cl-Ga	-1.5

Tab. 4.4: Operating parameters for hybrid 2B (CCG, High-E, rsA). Except of the Source potential, which is referenced to ground, all potentials are referenced to Source. The Drain potential of roughly 2 V is defined by the readout chip.

studied using the DAQ prototype system. The internal gain of the structures with a high-E implantation of $g_Q = 308 pA/e^-$ is in agreement with three dimensional simulations. It has been shown, that a complete clear of the structures is possible, albeit the required potential difference at the clear contact of around 14 V is higher than desirable.
The performance of the on-chip pedestal subtraction of the readout chip CURO 2 is satisfying, albeit the speed requirements of ILC are not yet reached. In the next chip iteration, the timing of the coarse sample cell must consider a decent time for a clear pulse. The timing of the output stage is too slow by orders of magnitude. Using a reasonable timing yields a homogenous pedestal distribution, demonstrating that effects of the sensor matrix are not observable. The comparator pre-charge, intended for a fast hit discrimination even of signals near the threshold, leads to high currents of several ten mA in a fast, pulsed operation. This mechanism has to be carefully designed in the next version of the chip to avoid fluctuations of the supply voltage. In the present iteration fluctuations has been observed leading to unstable thresholds.
The system noise is by far dominated by the sampling noise of the CURO, which is roughly ENC = $300e^-$. A signal to noise ratio of $S/N_{60\,keV} = 51$ has been shown using a ^{241}Am γ-source (59.5 keV).
The performance of the system in a beam test environment using a full readout and the on-chip zero-suppression is presented in the following chapters. Some of the effects like the corruption of the first rows or the unstable threshold were not observed until the detailed analysis of the test beam data.

5. Beam test studies

In this chapter the performance studies of the DEPFET ILC prototype system in a beam test environment are presented. After a description of the experimental setup different event reconstruction methods are discussed. The chapter deals with the non-zero-suppressed readout, which allows comprehensive studies of the system properties like signal, noise and charge sharing between the pixels. The spatial resolution is studied with different algorithms. In the following chapter the performance of the on-chip zero-suppression is being analyzed and compared to the full readout.

5.1 Experimental setup

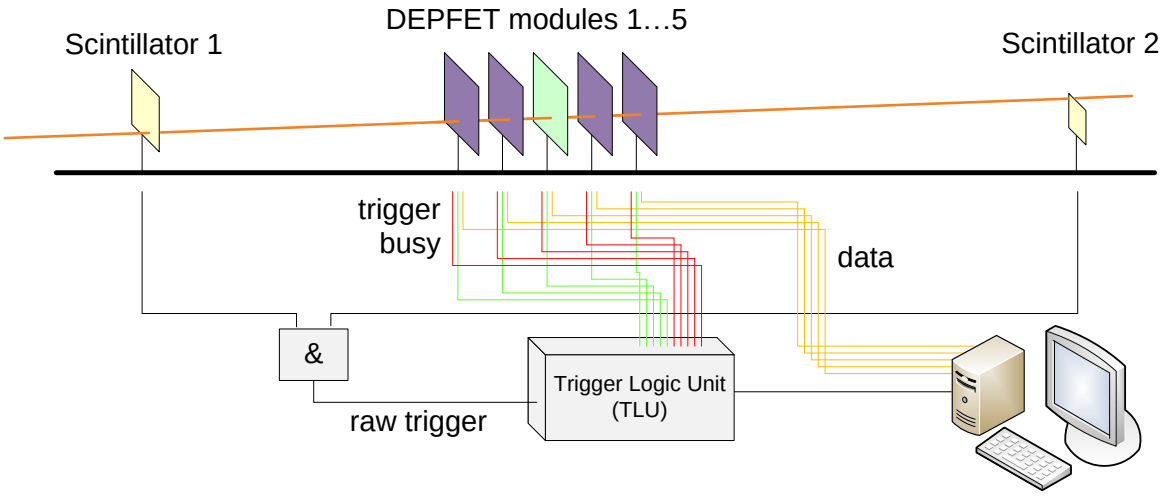


Fig. 5.1: Schematic drawing of the experimental beam test setup with five DEPFET sensors. The Trigger Logic Unit (TLU) receives the raw trigger signal (coincidence of two scintillator signals) and the TLU busy signals of each device. The trigger is passed to all devices simultaneously if none of the devices is busy.

The performance of different types of DEPFET matrices has been studied using a 6 GeV electron beam at DESY (Hamburg, Germany) and pion beams with an energy of up to 120 GeV at CERN H8 (Geneva, Switzerland). Most of the properties of the beam like particle contents, intensity and particle energy are well known and the beam is well collimated. The data presented here has been obtained at CERN in October 2006.
A schematic drawing of the experimental setup is shown in figure 5.1. The device under test (DUT) is placed in the center. Two planes of a reference system (beam telescope) at each side of the DUT are used to reconstruct the particle tracks. Since the system is

not self-triggered, an external signal is needed to trigger the data acquisition in case of a particle passing the sensitive volume of the detector. This trigger signal is provided by the coincidence of two scintillation counters, one located at the front of the telescope and one on the back. To minimize the amount of events with particle tracks outside of the sensitive area of the DUT ideally a scintillator of the same dimensions as the DUT ($2.3 \times 2.8\ mm^2$) is used. The smallest scintillator which was available had a size of $4.2 \times 4.2 mm^2$, which corresponds to an area three times the size of the DUT. As multiple scattering is negligible (see below) and the tracks can be considered parallel one small scintillator is sufficient and its position along the beam is arbitrary. In this case a coincidence of two scintillators is used only to exclude triggers caused by the photomultiplier noise.
As a central element of the setup a *trigger logic unit* (TLU) receives the *raw* trigger signal, which is the coincidence of the two scintillator signals. In addition the TLU receives busy signals of each device. Since for a proper event analysis the data of *all* devices are needed, a raw trigger is only accepted and passed to all devices simultaneously if none of them is busy. If at least one of the devices is busy (for example during the readout of the sensor or the data transfer), the event is being ignored. The used TLU was built for the ATLAS testbeams [38].
For high precision studies of the spatial resolution, the reference system has to provide tracks with an error on the predicted position smaller than the intrinsic resolution of the device under test. From the first beam test studies at DESY it was known that the intrinsic resolution of the DEPFET matrices is smaller than $\sim 5\ \mu m$. The precision of these studies was limited by multiple scattering and the intrinsic resolution of the BAT[1] telescope.

Multiple Coulomb Scattering A charged particle traversing any kind of matter is deflected by a series of small angle scatters. Most of this deflection is due to Coulomb scattering from nuclei. As shown in fig. 5.2, after having passed the material the particle will be shifted by some distance with respect to the point of incidence. Also the particle direction will change by some deflection angle Θ. The mean deflection angle is given by:

$$\Theta_0 = \frac{13.6\text{MeV}}{\beta cp} \cdot z \cdot \sqrt{\frac{x}{X_0}} \quad \left[1 + 0.038 \quad ln\left(\frac{x}{X_0}\right)\right] \tag{5.1}$$

Here, p, βc and z are the momentum, velocity and charge number of the incident particle, and x/X_0 is the thickness of the scattering material in radiation lengths, which is $X_0 =$ 9.36 cm for silicon. The shape of the angular distribution is roughly gaussian for small deflection angles.
Calculating the deflection angles for a silicon sensor of 450 μm thickness yields an rms value of $\Theta_0 = 0.13\ mrad$ in the case of 6 GeV electrons at DESY compared to $\Theta_0 = 6.3 \cdot 10^{-3}\ mrad$ for 120 GeV pions at CERN. Ignoring all other scattering materials like aluminum foil and air at DESY the average scattering of one plane already leads to a deviation of 3.2 μm in the next plane at a distance of 25 mm, which is too much for high precision position resolution studies of the DEPFET sensor. With a deviation of $\sim 0.16\ \mu m$ at CERN the multiple scattering is negligible. Therefore, further testbeams

[1] At DESY the Bonn ATLAS Telescope (BAT) has been used [38].

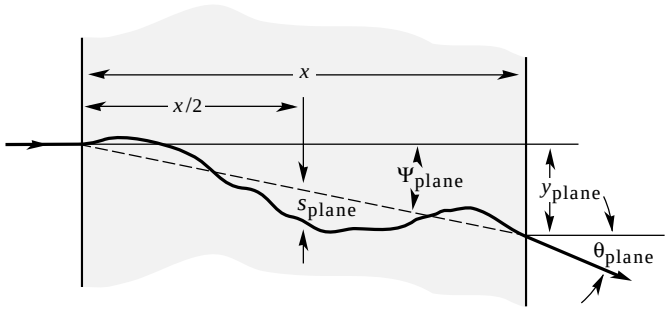


Figure 5.2: Principle of Coulomb multiple scattering. After a series of small angle scattering processes in the electrical field of the nuclei the particle direction changed by some deflection angle Θ (from [12]).

took place at the CERN SPS. Several DEPFET planes were operated simultaneously to build a stand-alone, high precision beam telescope (see fig. 5.3). This is discussed in more detail in sec. 5.7.

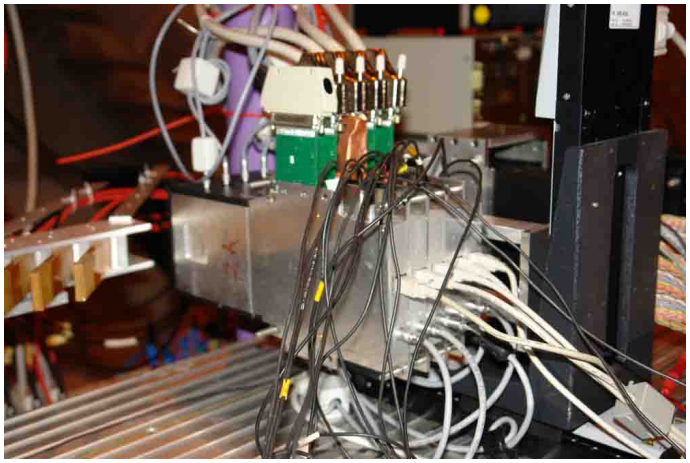


Figure 5.3: Photograph of the beam test setup. Five modules in aluminum housings are attached to each other and aligned in the beam using an x-y-motor stage. The sensors are covered only by thin aluminum foils.

The whole setup was mounted on a dedicated, robust aluminum table. For a minimal distance of the sensor planes the modules were attached to each other. This block of five modules has been aligned in the beam using a precise x-y motor stage. The drawback of this method is that an individual alignment of the modules is not possible. Due to the small sensor dimensions the overlap of all five modules turned out to be only 30% of the sensitive area, which translates to $1.6 \times 1.4\ mm^2$. This corresponds to only 11% of the scintillator area. Thus, a particle track is expected in every tenth event. For the next beam test period a smaller scintillator and individual motor stages for each module are planned for a minimization of empty events. The operating parameters of the DEPFET sensors were set to the values listed in table 4.4. The temperatures have been monitored using a temperature sensor attached to the hybrid close to the detector matrix.

5.2 Event reconstruction

The raw data of a non-zero-suppressed event contains signal amplitudes of all pixels of the matrix. The signal is composed of a pixel-specific offset (pedestal), random and correlated noise and possibly the particle signal. As already discussed in chapter 2.1.3, a typical MIP signal is spread over several pixels, a so-called *cluster*. To reconstruct the

event the cluster signal is extracted and a set of pulse height corrections is being applied. This is done off-line in different analysis steps.
The event reconstruction starts with a pedestal correction of the raw data. After a coarse hit exclusion the correlated noise (row-wise common mode noise) is being corrected for. In a second analysis iteration hits are identified using the noise values to define the hit cut. Dead, hot and edge pixels are masked and excluded from further processing. The cluster properties and the precise impact position are analyzed thereafter. In the following the determination of pedestals and noise are discussed, followed by a discussion of the correlated noise and the cluster reconstruction. The special treatment of zero-suppressed data is discussed in section 6.

5.2.1 Pedestals

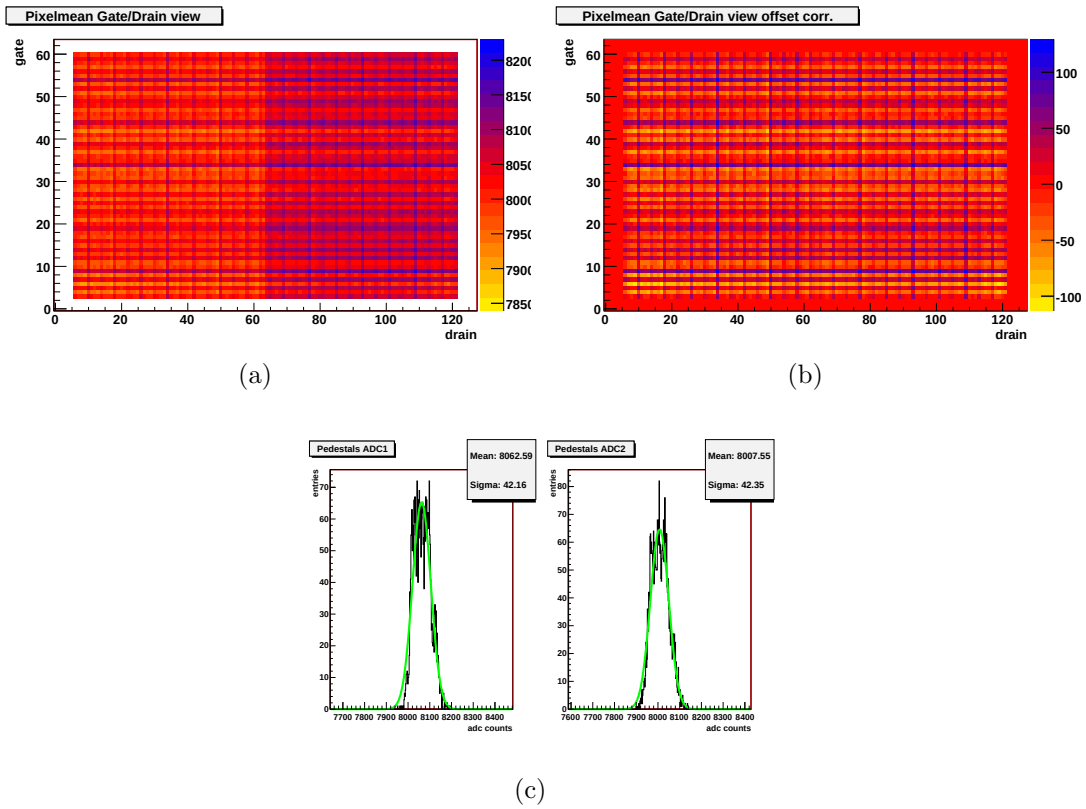
The mean pedestal value $\bar{p}_i$ for pixel i using n events is determined by calculating the mean value:

$$\bar{p}_i = \frac{1}{n} \cdot \sum_{j=0}^{n} p_i^j$$

It has been shown, that 3000 events of a run are sufficient for a pedestal calculation and that the pedestals are stable within a run of typically 30000 events[2]. A typical pedestal distribution of the matrix as found in the beam test data is shown in fig. 5.4. After correcting for the offset of the two ADCs, the dispersion throughout the matrix is found to be larger than 40 ADU (rms). Thus, an off-line pedestal correction is necessary, although the matrix pedestals are subtracted on-chip already. The main contributions are offsets of the channels of the readout chip, remnants of the on-chip pedestal subtraction and oscillations at the output stage. As discussed in chapter 4.4, the effects of the oscillations, particularly the regular pattern, can be avoided by a longer duration of the output sampling.
In zero-suppressed mode pedestals are - by definition - not directly measured. Pedestals of non-zero-suppressed data can be used for a pulse height correction, provided the pedestal distributions do not change. Therefore, the stability and reproducibility of the pedestals has been studied. An exemplary comparison of the pedestal distributions of different runs using the same matrix and identical settings is shown in fig. 5.5. Both pedestal distributions are pixel-wise subtracted from each other. This method has been used to compare 22 non-zero-suppressed runs with identical settings (figure 5.6). An average deviation from an arbitrary reference run of below 10 ADU has been observed.

In conclusion, the distribution is stable within a run but changes from run to run. Maximal deviations of around ± 30 *ADU* require an individual pedestal determination for each run, since the cut for the neighboring pixels is in the same order of magnitude (see next section). A pulse height correction of zero-suppressed data, where no off-line cuts are applied, is possible with any (or an average) pedestal pattern derived from non-zero-suppressed data. The alternative is a common offset, which may yield deviations in the order of 40 *ADU* (rms).

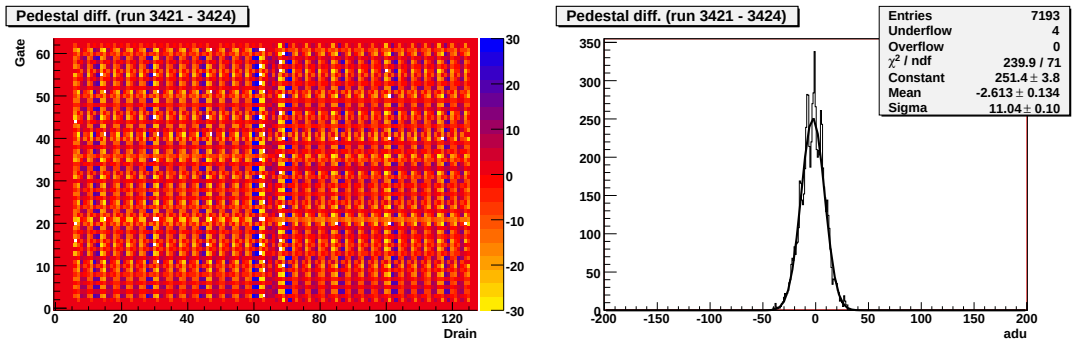
[2] 30000 events were obtained within approximately 30 minutes.



(a) (b)

(c)

Fig. 5.4: Pedestal distributions of hybrid 2B (beam test data, default settings). In fig. a) an offset of both ADCs (left and right half of the event display) is clearly visible, the pedestal value is plotted color-coded in ADC units. This offset is corrected in b). The dispersion (rms error) is found to be in the order of 40 ADU (fig. c). The regular pattern originates from the output stage of the readout chip, most likely due to crosstalk.



(a) (b)

Fig. 5.5: Comparison of the pedestal distributions of two different non-zero-suppressed runs using the same matrix and identical settings. The pedestal values are subtracted pixel-wise.

For the pedestal determination ideally a set of data is used, which is acquired in absence of

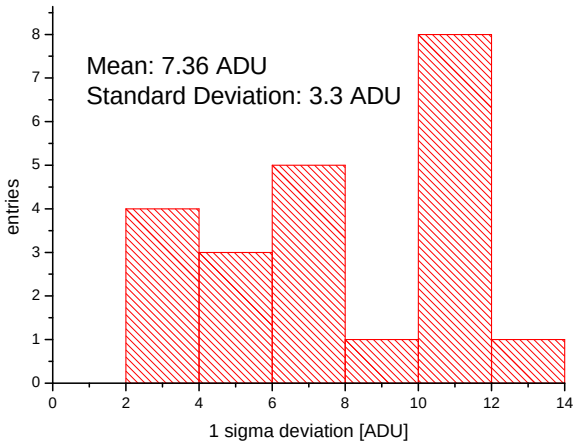
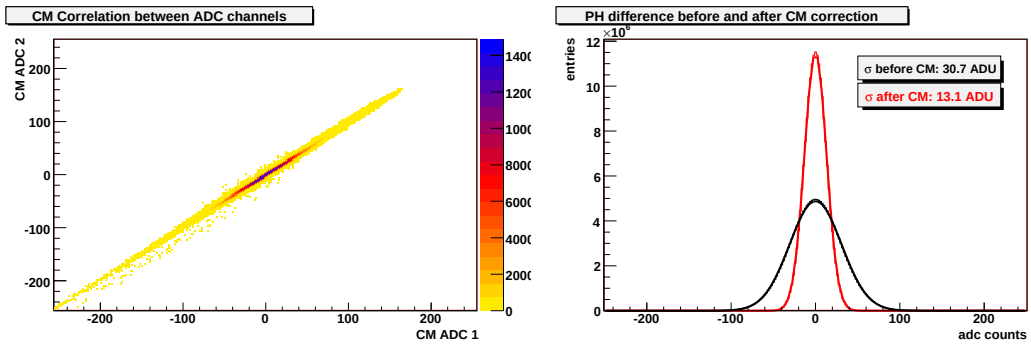


Figure 5.6: A comparison of 22 non-zero-suppressed runs (using identical settings) with a single reference run has been done similar to fig. 5.5. The standard deviations derived from those distributions are plotted here. The average pedestal offset was found to be 2.6 ADU (not shown), while the average rms deviation was found to be 7.4 ADU.

beam. Since this is unfortunately not practicable for each run, identifying and excluding the hits from the data sample is necessary. The occupancy is found to be low and the error resulting from particle hits included in the pedestal sample is therefore small ($< 10^{-3}$). An option is to fit a gaussian to the signal distribution of a channel and use the mean value of the fit rather than the arithmetic mean value. In this case, few entries with large pulse heights are not taken into account.

5.2.2 Noise



(a) (b)

Fig. 5.7: (a) Correlation of the row-wise common mode noise values of the two ADC channels. (b) Difference of the signal height distribution before and after common mode noise correction.

The noise has two components, a random variation and a correlated variation, the common mode noise. The latter may be corrected for. The common mode noise originates from low-frequency baseline fluctuations. This may be RF pickup in the matrix, readout chip or PCB as well as variations of the supply voltages. If and which kind of common mode noise is present in the data is analyzed by pulse height correlations between all channels of the matrix. In the present case correlation plots reveal row-wise fluctuations. This is not unexpected, since the signals of a row are sampled simultaneously and the time span

between two consecutive rows is much larger than the sampling time itself. Consequently, the common mode noise is corrected for by determining the average pulse height of all valid pixels in a row of double pixels. This is done after the pedestal correction and hit rejection. Thus, the gaussian common mode noise distribution is centered around zero. A hit exclusion is necessary during common mode correction, as a hit signal within a row of only 128 pixels would significantly distort the common mode value. Therefore, all pixels carrying a signal larger than typically $4 \cdot \sigma_{noisepeak}$ and all adjacent pixels (potentially carrying also parts of the signal) are excluded from the analysis. A correlation of the common mode values of the two readout chains (ADCs 1 and 2) is shown in fig. 5.7(a), demonstrating that the common mode noise is originating from an earlier stage of the readout chain, before the two chains are separated, which happens inside the readout chip. Figure 5.7(b) shows the difference of the integrated pulse height distributions of a run before and after the common mode correction. The standard deviation of the distribution is reduced from ~ 30 *ADU* down to $12 - 15$ *ADU*. The exact origin of the common mode noise is not identified yet, but various studies indicate that the readout chip is the dominant source.

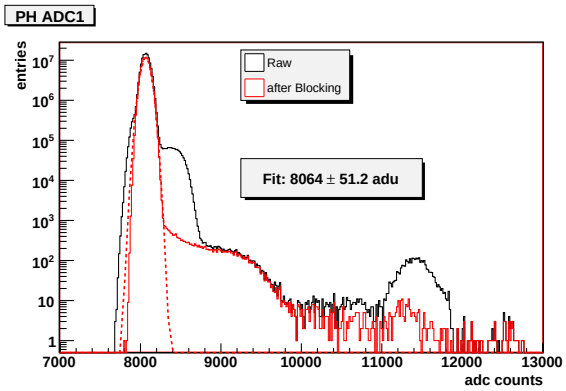


Figure 5.8: Raw signal distribution before and after blocking of bad pixels. The deviations from the gaussian shape of the noise peak and the peak at 11500 ADU (before blocking) are due to readout artifacts (see chapter 4.4).

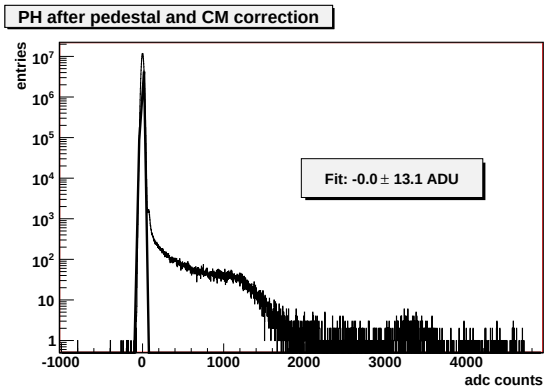
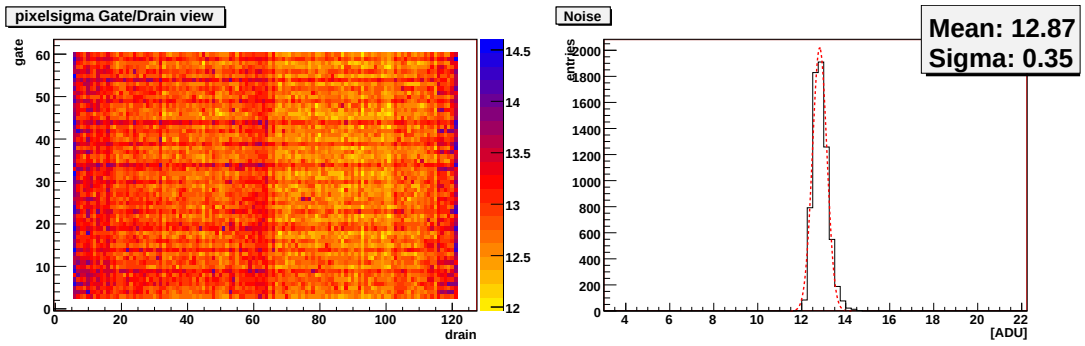


Figure 5.9: Signal distribution after all corrections. The noise peak is centered around zero. Unclustered signal entries are mainly below 2000 ADU.

The pulse height distribution of a typical beam test run before (raw) and after all corrections is shown in figures 5.8 and 5.9. Corrections include the pedestal subtraction, common mode noise correction and masking of bad pixels, particularly the first read out rows of each event (see chapter 4.4). The noise distribution after hit exclusion is shown in fig. 5.10. An average noise per channel is found to be (12.9 ± 0.4) ADU, which translates to (318.3 ± 7.4) e^- based on the calibration presented in chapter 4.



(a) (b)

Fig. 5.10: Pixel-wise noise distribution. The noise is homogeneously distributed over the matrix with an average value of (12.9 ± 0.4) ADU (translates to (318.3 ± 7.4) e^-). The edge pixels are masked because of their higher noise figure.

5.3 Cluster reconstruction

The first step of the cluster reconstruction is the identification of the pixel that carries the highest signal that is larger than a certain threshold. This pixel is referred to as *seed*, and the respective threshold is called *seed threshold* (or cut). While the high signal of the seed allows a distinct separation from the noise, the extraction of the remaining charge in the surrounding pixels is usually a trade-off between a complete but noisy signal and a potentially incomplete low-noise signal. There is no general rule on how to do an optimal cluster reconstruction; it depends on the sensor type, the S/N and the applied position reconstruction algorithm. There are mainly three approaches:

- Fixed area: Integrate the charge of a fixed area around the seed pixel (for instance only the direct neighbors, i.e. 3×3 pixels). Easy, but subject to errors (depending on the size of the charge cloud and the pixel pitch either noise hits or too less pixels included).
- Fixed cluster size: Neighbors of the seed pixel are added in order of decreasing signal height until the desired cluster size is reached. For example, the η algorithm for position reconstruction uses the two highest pixels in each direction.
- Variable cluster size: Add all neighbors of the seed pixel which carry a signal larger than a certain threshold. This so-called neighbor cut (or next cut) can be chosen smaller than the seed threshold.

In this thesis, mainly the latter method of a variable cluster size is presented and studied with different cuts. In case the main charge deposition is located in a the masked pixel, the neighboring pixels carrying only part of the charge must not be falsely identified as seed. This is assured by also masking the neighboring pixels, though for seed finding only. As particle signals are intrinsically stored within a DEPFET pixel until it is read out, the probability for multiple hits within a readout cycle is high. Hence, multiple seeds are allowed per event.

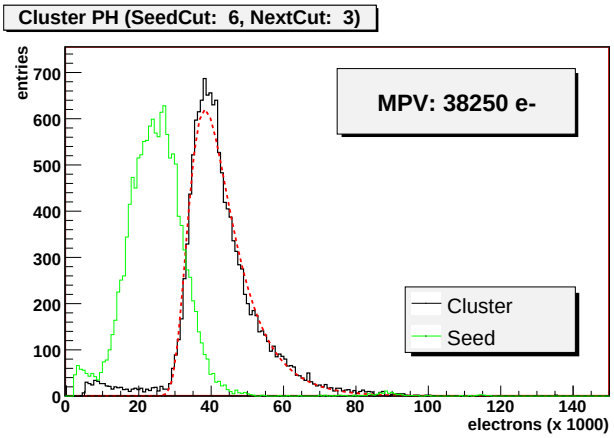


Figure 5.11: Cluster charge distribution. The cluster entries with a signal below the low signal tail of the Landau distribution are due to readout artifacts (see text).

The charge deposition is described by the Landau distribution (see chapter 2.1.1). The most probable energy loss (MPV) of (38250 ± 50) e^- (for a neighbor cut of $3\sigma_{noisepeak}$), as shown in figure 5.11, yields a S/N-ratio of 121 ± 3. The clusters were found to be homogeneously distributed over the sensor, no dead areas were observed. Also shown is the seed signal distribution which is used to determine a reasonable seed cut. With minimum seed signals of ≈ 8000 e^- signal and noise are clearly separated. Assuming a gaussian noise figure, with a cut of $5\sigma_{noisepeak}$ above the mean value, for instance, less than $3 \cdot 10^{-5}$ % of noise signals are falsely identified as hit. This corresponds to 1 pixel out of 50000 events. However, it turned out, that readout artifacts lead to signal entries between the nicely gaussian shaped noise peak and the low signal tail of the Landau distribution. These hits are rare compared to the amount of real hits (< 5 %), but are not observed in the absence of the beam. The low signal entries are not correlated to real particle tracks, as will be shown in section 5.8.

5.4 Charge sharing

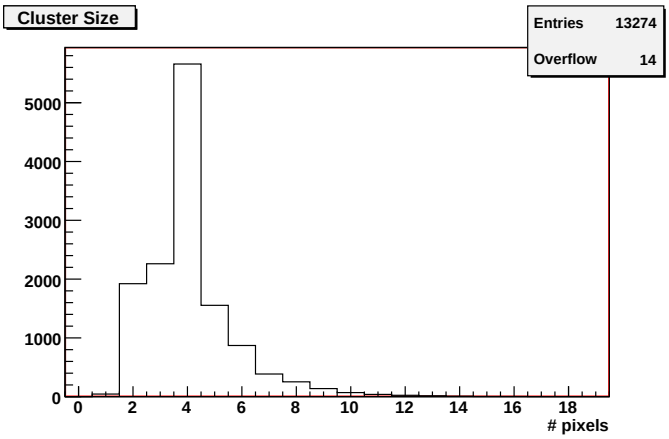


Figure 5.12: Cluster size distribution for a seed cut of $6\sigma_{noisepeak}$ and a neighbor cut of $3\sigma_{noisepeak}$. Mainly clusters of four pixels are observed.

In the case of a pixel pitch smaller than the dimensions of the charge cloud, the charge is shared by several pixels. Compared to a single pixel hit, where the impact position is always reconstructed at the pixel center, charge sharing can be exploited for a significantly

better position reconstruction. Given the calculated diameter of the charge cloud of $d \approx 35.6\mu m$ ($\pm 3\sigma$; see sec. 2.1.3) and the pixel pitch of the matrix of 36×22 μm^2 for geometrical reasons mostly cluster sizes of 3-4 pixels are expected. The observed cluster size distribution for a seed cut of $6\sigma_{noisepeak}$ and a neighbor cut of $3\sigma_{noisepeak}$ is shown in fig. 5.12. Mainly clusters consisting of four pixels are found, which is in agreement with the expectations.

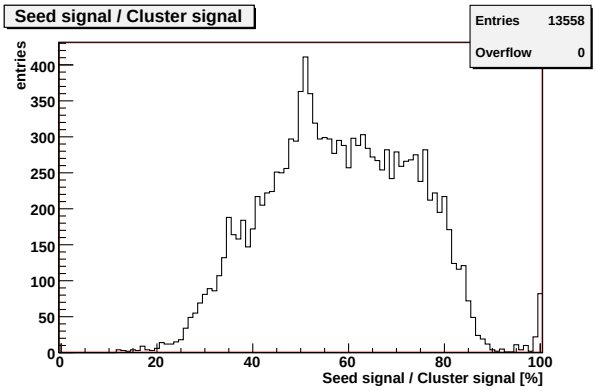


Figure 5.13: Ratio of seed and cluster signal.

The ratio between seed signal and cluster signal is shown in figure 5.13. Typically between 25 % and 85 % of the cluster charge is collected in the seed pixel.

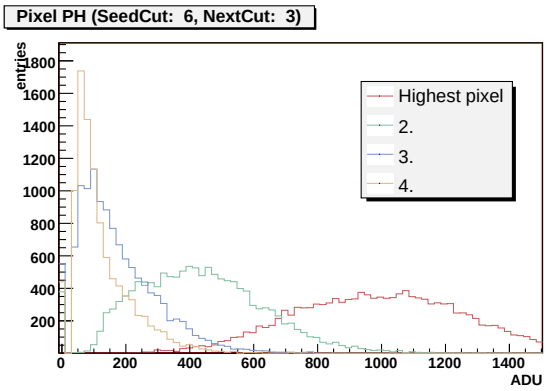


Figure 5.14: Charge distributions of the neighboring pixels to a seed pixel.

In figure 5.14 the charge distributions for the neighboring pixels are shown. It can be seen, that the highest neighbor is always larger than $5\sigma_{noisepeak}$, whereas already the second highest neighbor is not clearly separated from the noise. As the probability of charge deposition in the neighboring pixels is high, the cluster reconstruction can be improved by moderately lowering the neighbor cut. With a cut of $3\sigma_{noisepeak}$ the noise fake rate rises to 0.15%, which is still tolerable.

5.5 Track reconstruction

The reconstruction of tracks from the set of 2-dimensional hit information of the beam telescope starts with the precise alignment of the detector planes. One plane is used to define the zero-position. All other planes are iteratively aligned in terms of spatial coordinate and angles in all rotational axes by minimizing the deviation of a measured

hit from a straight line fit of the corresponding track. For the alignment procedure only events with a single hit in each plane are used. The alignment of the DUT is done in the same way.
After the alignment the tracking is done also by straight line fits. To derive an unbiased prediction of the hit position in the DUT plane, the DUT itself is not included in the fit. Furthermore, only tracks passing all planes are considered. In case of several hits in one or more planes, the tracks are selected by a χ^2 cut. The result of the fit is then extrapolated to the DUT plane which yields a prediction of a hit position in the DUT. The uncertainty of the prediction has to be much smaller than the expected intrinsic resolution of the DUT. The alignment and track reconstruction procedures are discussed in more detail in [27] and [9].
Assuming identical intrinsic resolutions and the absence of multiple scattering, the error on the predicted position is smallest in the center of the setup. Therefore, in the following the results of hybrid 5 are discussed, which was located at the center position in the beam test setup. The layout and the operating parameters are identical to hybrid 2B, which was used for all other studies and for the zero-suppression studies discussed in the following chapter. Unfortunately, the latter was one of the outer modules, with the largest tracking uncertainty. For the determination of the spatial reconstruction of the hits in the telescope planes the η-algorithm is used (see below).

5.6 Spatial resolution

The spatial resolution is commonly defined as the width of the residual distribution of predicted and measured hit position. While this number still includes the contributions of noise and the uncertainty of the predicted position, the corrected value is referred to as intrinsic resolution of the device under test. Different reconstruction algorithms are discussed, the binary reconstruction, the Center of Gravity and the η-algorithm.

5.6.1 Binary reconstruction

With a binary reconstruction the impact position is reconstructed to the center of the pixel carrying the highest signal. The theoretically achievable spatial resolution in absence of noise depends only from the pixel pitch p:

$$\Delta x = \sqrt{\langle (x - x_{rec})^2 \rangle} = \sqrt{\frac{1}{p}\int_0^p \left(x - \frac{p}{2}\right)^2 dx} = \frac{p}{\sqrt{12}} \tag{5.2}$$

In the case of the studied DEPFET sensor this translates to $\Delta x = 10.4\ \mu m$ and $\Delta y = 6.4\ \mu m$ respectively.

5.6.2 Center of Gravity

With the Center of Gravity (CoG) method the impact position is reconstructed as the charge weighted average pixel position independently for both directions x and y:

$$x_{CoG} = \sum_{cluster} P_i x_i / \sum_{cluster} P_i \tag{5.3}$$

with P_i the signal of pixel i. This method assumes perfect linear charge sharing and a box-shaped charge cloud. In reality, a gaussian charge cloud is found, leading to a systematic error on the reconstructed position towards the pixel center (see illustration 5.15). Additional errors are made by cuts and noise. Inhomogeneous pixel response especially in the inter-pixel areas is not taken into account. As shown in fig. 5.16 residuals of $\sigma_x = 5.40\ \mu m$ and $\sigma_y = 2.13\ \mu m$ have been observed. Note, that these values includes the error on the predicted position, which is estimated in chapter 5.7.

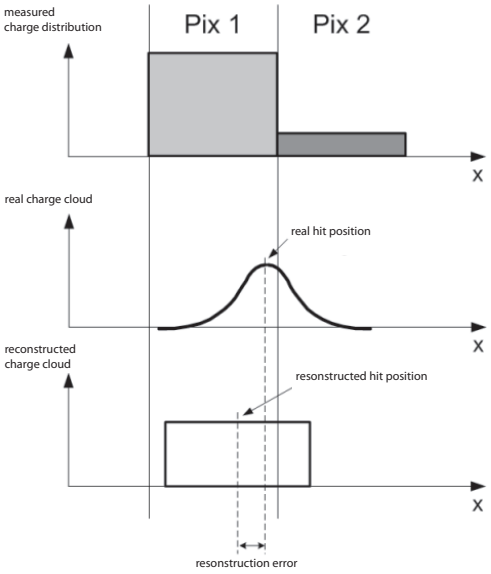
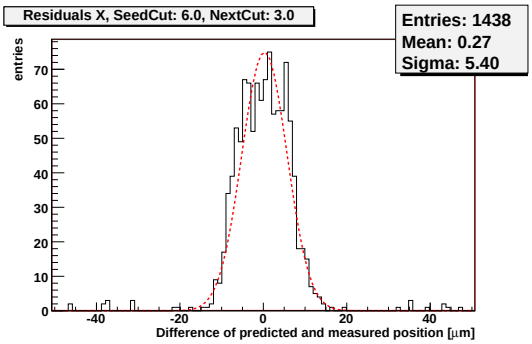
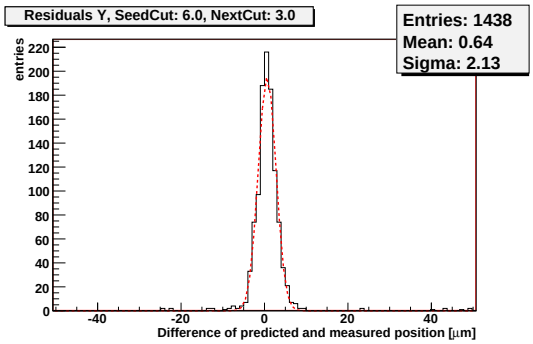


Figure 5.15: Problem of the linear reconstruction. A box shaped interpretation (bottom) of the observed charge distribution (top) leads to a systematic error of the reconstruction, as the actual gaussian shape of the charge distribution is not taken into account (from [18]).



(a) CoG X, Seed cut 6σ.



(b) CoG Y, Seed cut 6σ.

Fig. 5.16: Residual distribution (CoG, Mod 5, DUT in the center).

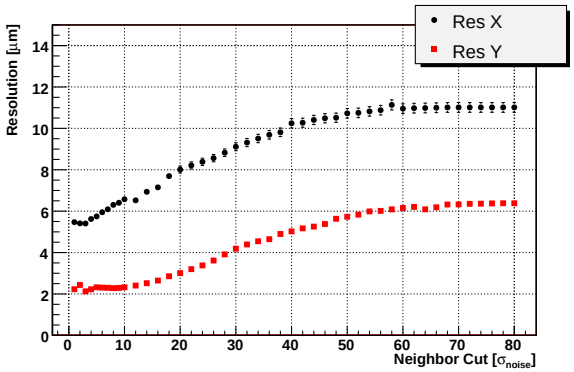


Figure 5.17: Spatial residuals as a function of the neighbor cut (CoG, Mod 5, DUT in the center). For very high cut values only the seed pixel is used for spatial reconstruction, leading to a binary reconstruction.

The impact of the noise has been studied by varying the cut neighbor pixels. The result is shown in fig. 5.17. In x-direction (wider pitch) the best resolution is found at a neighbor cut of $2\sigma_{noisepeak}$. In y-direction the best resolution is achieved in the range of $3\sigma_{noisepeak}$ to $10\sigma_{noisepeak}$, demonstrating an optimal ratio of charge cloud and pixel pitch. Lowering the cut increases the impact of the noise and worsens the resolution. At the limit of very high neighbor thresholds only the seed is used for the event reconstruction, which is equivalent to the case of binary reconstruction. Here, the result is in agreement with the theoretical limits stated above. Note, that the theoretical limit ignores the system noise and tracking uncertainties.
Figure 5.18 shows the probability of multi-pixel clusters for both directions. As charge sharing between at least two neighboring pixels can be exploited for a significant improvement of the position reconstruction, the neighbor threshold can be optimized in that respect. While in the case of a reasonable cut of $3\sigma_{noisepeak}$ the probability for at least one neighbor is 98 % in y-direction, it is still 80 % in the wider x-direction.

5.6.3 η algorithm

The so-called η-algorithm is a widely used reconstruction algorithm. The reconstruction function is easy to determine and it is independent of the shape of the charge cloud.
The variable η describes the charge sharing between two pixels, individually for both directions of a matrix:

$$\eta_x = \frac{Q_{left}}{Q_{left} + Q_{right}} \tag{5.4}$$

where Q is the signal charge in the respective pixel. Only the two pixels carrying the highest signals are taken into account. The distribution of η is determined from a sufficiently large data set of hits uniformly distributed over the pixel, and is specific for the combination of pixel geometry, charge sharing and S/N.
Assuming that the detection efficiency is uniform over the pixel, the impact position is reconstructed by

$$x_{rec} = x_{right} - p_x F(\eta_x) \tag{5.5}$$

Here, x_{right} is the position of the right pixel and p_x the pixel pitch in x-direction. The reconstruction function $F(\eta)$ is then defined by

$$F(\eta_x) = \frac{1}{N_0} \int_0^{\eta} \frac{dN}{d\eta'} d\eta' \tag{5.6}$$

where $dN/d\eta$ is the η distribution. The observed η distributions and the respective reconstruction functions of hybrid 5 are shown in fig. 5.19. The less charge sharing and noise are present the more pronounced are the peaks of the η distribution close to 0 and 1. In the y-direction a much larger charge sharing is observed than in the wider x-direction. While in the first case, the charge sharing is too uniform for a real improvement of the reconstructed position, in x-direction the η-algorithm is much more powerful and corrects for the non-linear charge sharing. One reason for the uniform distribution in y-direction is a geometrical one, i.e. the ratio between charge cloud and pixel pitch, but there might also be some other mechanism that facilitates the interaction between neighboring pixels in that direction. While the 'clear' implantation between neighboring pixels in the x-direction assures a well defined separation, the borders in the y-direction are not so distinct. In addition, due to the double pixel structure the pixel transitions are not symmetric (see figure 2.13). This important issue of in-pixel charge collection efficiency and charge sharing is currently under study [31].
The spatial residuals achieved with the η reconstruction are shown in fig. 5.20. While the improvement compared to the CoG-method in direction of the smaller pixel pitch is below 20% ($\sigma_y = (1.78 \pm 0.01)\mu m$), in x-direction the reconstruction error goes down to $\sigma_x = (3.44 \pm 0.03)\mu m$, which is an improvement of 39%. Note, that these numbers include

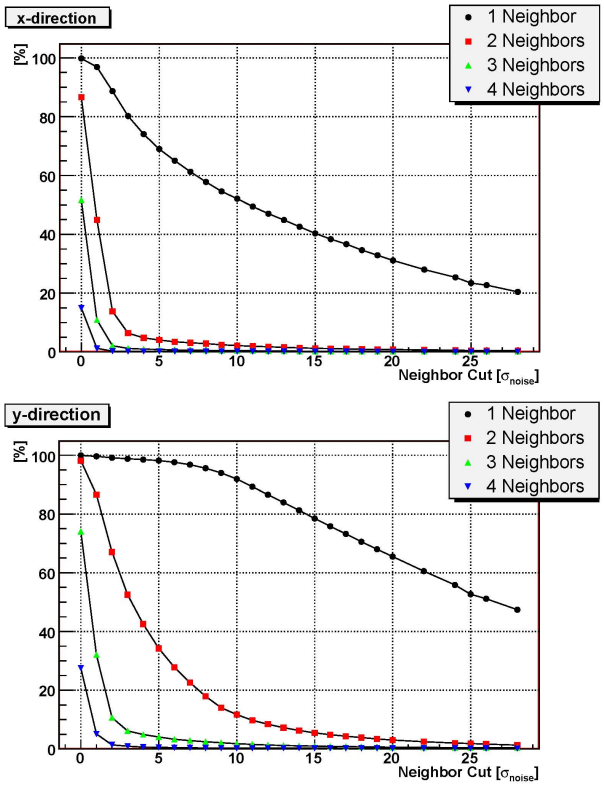
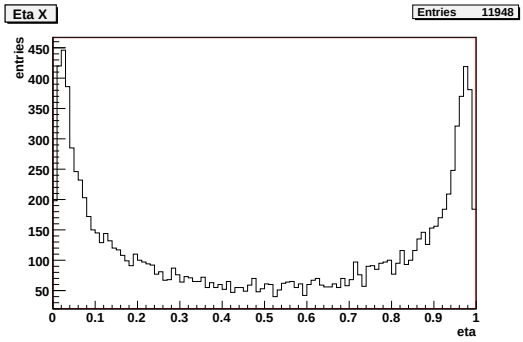
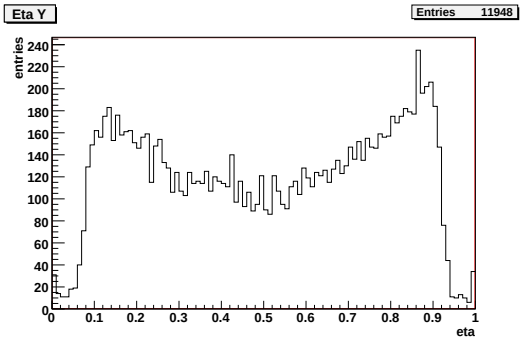


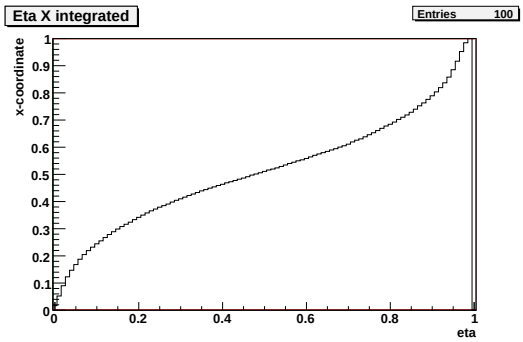
Figure 5.18: Probability of charge sharing for a pixel pitch of 36 μm in x-direction (upper plot) and 22 μm in y-direction (lower plot). For a $3\sigma_{noisepeak}$ neighbor cut the probability for a charge deposition in at least one neighboring pixel is 98 % in y-direction and 80 % in the wider x-direction.



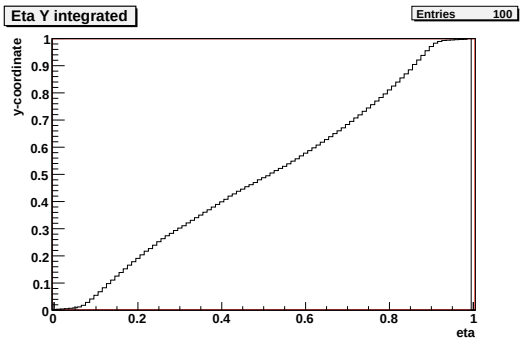
(a) η distribution in x-direction.



(b) η distribution in y-direction.

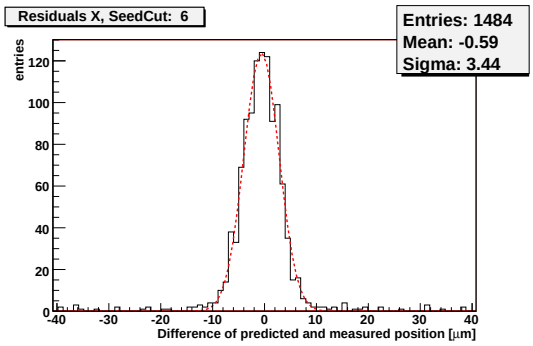


(c) $F(\eta)$ in x-direction.

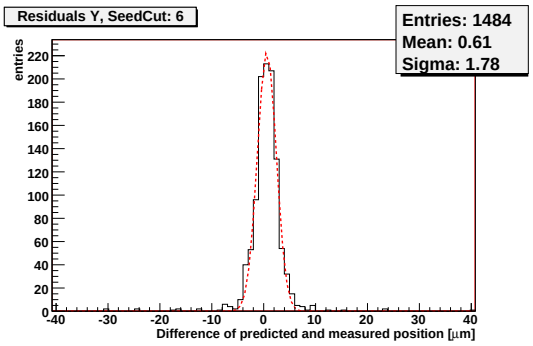


(d) $F(\eta)$ in y-direction.

Fig. 5.19: η distributions of Mod 5 (DUT in the center) and the respective reconstruction functions $F(\eta)$.



(a) Residual distribution in x-direction.



(b) Residual distribution in y-direction.

Fig. 5.20: Spatial residuals using the η reconstruction algorithm (Mod 5, DUT in the center, seed cut 6σ).

also the position extrapolation error of the telescope.

5.7 DEPFET beam telescope

With five modules a fully functional, high resolution DEPFET beam telescope has been operated. For a simultaneous operation of several DEPFET modules the DAQ software had to be extended. A new top layer multi-module Run Control application has been implemented, combining the administration and data acquisition functionality. Due to its modularity no principal modifications of the DAQ structure were necessary. The software can handle the operation of five modules; an extension to a larger number is straight forward. All modules are processed in separate applications, allowing for a parallel signal processing. The software package has been optimized with a dedicated timing and bottleneck analysis tool.
The dead time during the readout of the on-board buffered data is reduced to the pure data transfer via the USB 2.0 link, which takes $25ms$ every 16 frames. All further processing steps are done after re-enabling the data acquisition. This leads to a maximum readout rate comparable to the readout of a single module. During the beam test, the rolling shutter readout mode was used with a frame readout duration of $896\mu s$. This yields a maximum continuous readout rate of 254 Hz. During the readout of a frame no further triggers are accepted. However, in case of a high particle flux the integrating character of the DEPFET devices leads to many events with multiple hits. As discussed in more detail in the next section (5.8) the detection of all later particles passing the detector during the readout of a frame depends on the impact position in each plane and the respective readout progress. Since this complicates efficiency and purity studies, a modification of the readout sequence with an extendable readout may be beneficial for the next beam test period.

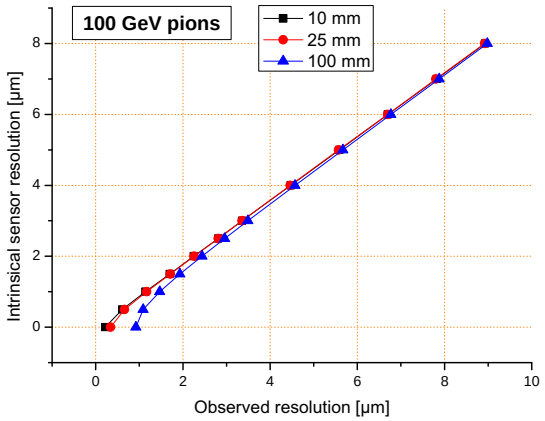


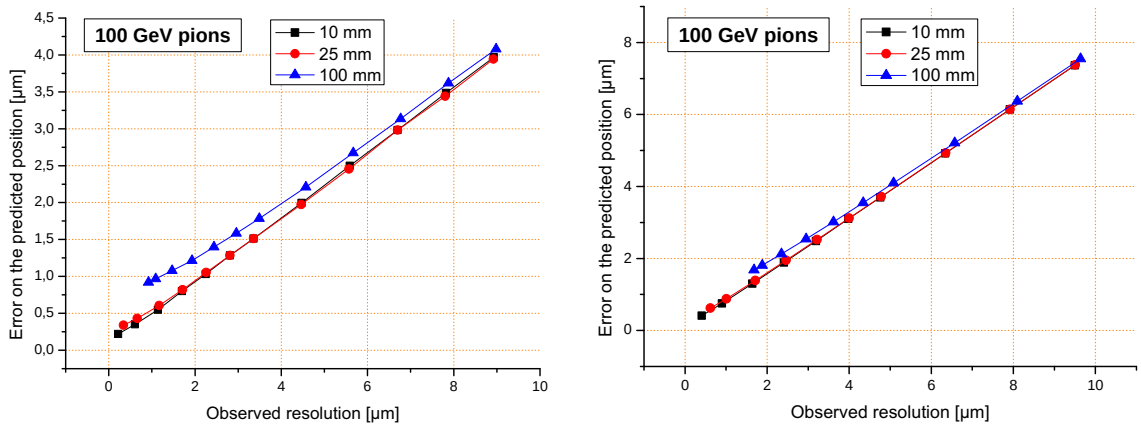
Figure 5.21: Intrinsic sensor resolution as a function of the observed resolution for different telescope plane spacings (Geant 4 simulation). The DUT is placed in the center (TT D TT).

The precision of the telescope has been simulated for different telescope configurations [8]. The basis of the Geant 4 simulation is a set of five identical planes (four telescope planes and one DUT in the center) of silicon detectors with a thickness of $450\mu m$ each and a variable distance. Tracks of pions are generated considering the energy dependent multiple scattering. The impact positions of the five planes are reconstructed with a variable intrinsic resolution, which is the same for all planes. A linear fit through the reconstructed positions of the four telescope planes is used to determine the predicted position on the DUT plane. The observed resolution (width of the observed residual

distribution) is a convolution of the intrinsic resolution of the detector and the uncertainty of the predicted position:

$$\Delta^2_{observed} = \Delta^2_{intrinsic} + \Delta^2_{telescope} \tag{5.7}$$

The intrinsic resolution depends not only on the S/N, the charge cloud and the pixel pitch of the detector, but also the errors of the spatial reconstruction are included. Please note, that the simulation ignores alignment errors.



(a) DUT placed in the center.

(b) DUT placed outside the telescope.

Fig. 5.22: Uncertainty of the predicted position in the DUT plane as a function of the observed resolution for different telescope plane spacings (Geant 4 simulation). A sub micron precision is reached with a TT D TT configuration using a plane distance of 25 mm if the observed residuals of all telescope planes are below $2\mu m$.

Figure 5.21 shows the observed resolution as a function of the intrinsic resolution of the sensors. In turn, the uncertainty of the predicted position as a function of the observed resolution is shown in figure 5.22. It can be seen, that a sub micron precision of the telescope is reached with a plane distance of 25 mm if the observed residuals of all telescope planes are below $2\mu m$. For the moment, this has not yet been achieved. In environments with negligible multiple scattering like at the CERN testbeam facilities, the lowest tracking uncertainties are reached for a DUT placed in the center of the setup with an equal distance to the two telescope planes at both sides each. This changes if multiple scattering must be considered. In such cases the DUT must be placed as close as possible to one of the detector planes. This is discussed in more detail in [27].
Based on the simulations the intrinsic resolution of the sensor is calculated using equation 5.7:

$$\Delta x_{intrinsic} = 3.0\mu m \tag{5.8}$$

$$\Delta y_{intrinsic} = 1.5\mu m \tag{5.9}$$

In figure 5.23 the uncertainty of the predicted position is compared for two different positions of the DUT: in the center between the telescope planes (TT D TT) and as

one of the outer planes (TTTT D). It can be seen, that the tracking performance is significantly better in the first case. The zero-suppression (discussed in the next chapter) has been studied with one of the outer modules.

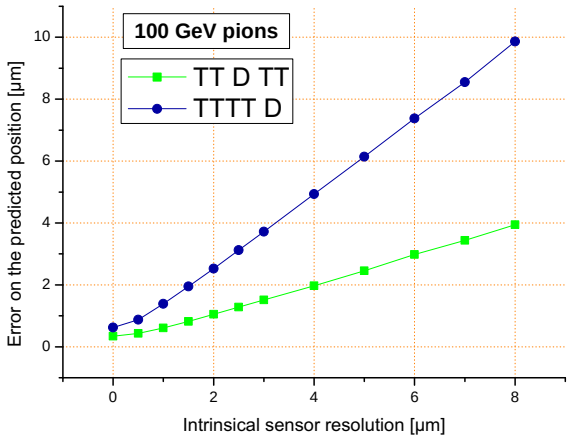


Figure 5.23: Comparison of the uncertainty of the predicted position as a function of the intrinsic sensor resolution for two different positions of the DUT (DUT in the center: TT D TT, DUT outside: TTTT D). The plane distance is 25 mm and the beam energy is 100 GeV. The zero-suppression has been studied with one of the outer modules.

5.8 Detection efficiency

For a vertex detector a detection efficiency close to 100% is required, which means, that ideally every particle passing the detector has to be detected. The efficiency is defined as

$$\varepsilon = \frac{\text{no. of detected hits}}{\text{no. of predicted hits}} \tag{5.10}$$

where only hits which correspond to predicted hits are taken into account. Given the high signal to noise ratio of the DEPFET sensor an efficiency close to 100% is expected. However, due to the periodical clearing of the internal gates the current row-wise readout scheme with on-chip pedestal subtraction comprises a short dead-time during which the particle signal is not detected. For an envisaged row rate of 20 MHz two current samples (signal plus pedestal first and only pedestal thereafter) and a clear have to be performed within 50 ns, for simplicity in the following a clear pulse of 10 ns and current samples of 20 ns each are assumed. Depending on the exact time of the particle passage through the row currently being processed three situations have to be distinguished, see figure 5.24:

1. Particle passage between t_0 and t_1: If the signal develops during the first sampling, the signal is probably not settled sufficiently. The result is a too low signal.

2. Particle passage between t_1 and t_2: The signal is cleared and thus lost.

3. Particle passage between t_2 and t_3: If the sampling is long enough to account for the signal, a particle passage during the pedestal sample will result in a negative signal amplitude, since in this case signal plus pedestal current are subtracted from the pedestal current.

These inefficiencies are not avoidable using the foreseen readout scheme. The probability of a particle passage through the row currently being processed is $p = 1/n_{rows} = 0.1\%$. For a 100% dead-time free readout in principal a readout scheme is possible which foresees a clear during the bunch pauses only and a repetitive readout during the spill. While also a higher row rate is achievable, the signal offset continuously increases due to leakage current. This complicates in particular the zero-suppression.

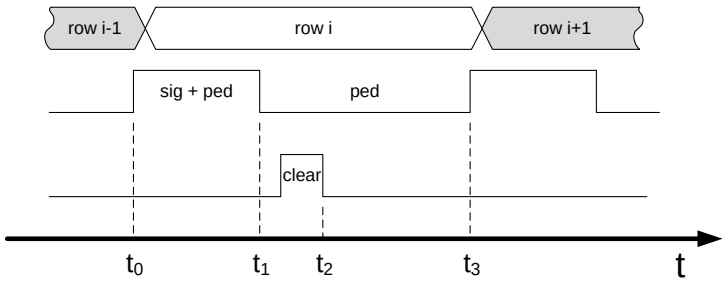


Figure 5.24: Signal processing using a row-wise readout with on-chip pedestal subtraction. In case a particle hits the row currently being processed, this hit might not be detected.

The present system is operated at a lower row-rate. As long as the system is ready and no trigger occurs, the matrix is cleared row-wise without being read out. The critical phase of the sequence is shown in figure 5.25. During the time between trigger query and the end of the clear pulse, which takes $t_1 - t_0 = 140\ ns$ using the beam test sequence, the particular row is insensitive for particles. The processing time of a row is $t_2 - t_0 = 200\ ns$. Thus, the probability for a particle hit in that time window is $p = 140\ ns/12.8\ \mu s = 1.1\%$.

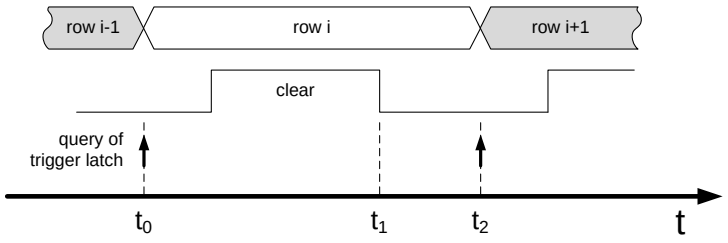


Figure 5.25: Row-wise clear of the present system operated in the beam test. As long as the system is ready and no trigger occurs, the matrix is cleared row-wise without being read out. Hits in the row currently being processed are lost, if the impact time is between t_0 and t_1.

In the following the beam test results are discussed. As will be shown, inefficiencies due to the dead time could unfortunately not be distinguished from system related inefficiencies with the used beam test setup. The efficiency study has been done for Hybrid 2B using 3020 tracks within the region of interest, which is the overlap region of all detectors including the DUT and the trigger scintillators. These tracks were reconstructed from a sample of 90000 events. In the DUT hits are looked for in an area of ± 2 pixels from the predicted position.

At this point some of the terms used in the following discussion need to be declared. A *particle* denotes a physical event. A *hit* is a signal detected with one of the sensors which is larger than the used cut, ideally due to a traversing particle. If hits in all telescope planes are found a *track* is reconstructed, the hits are then referred to as *track points*. The track extrapolation to the DUT plane defines a hit *prediction* (or predicted hit).

The efficiency as a function of the cluster cut is shown in figure 5.26. If all tracks within the region of interest are taken into account an efficiency of $\varepsilon = 89.8\%$ is observed for a cluster cut of up to 1100 ADU. This corresponds to a number of 308 particles which were not detected. A detailed analysis of the inefficiencies revealed three issues, all related to the system readout rather than to the sensor itself. In the following these issues are discussed and a plausibility check is presented.

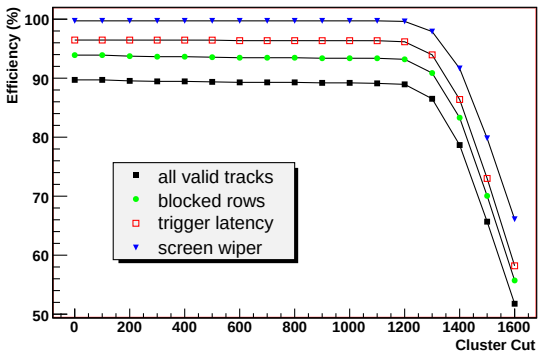
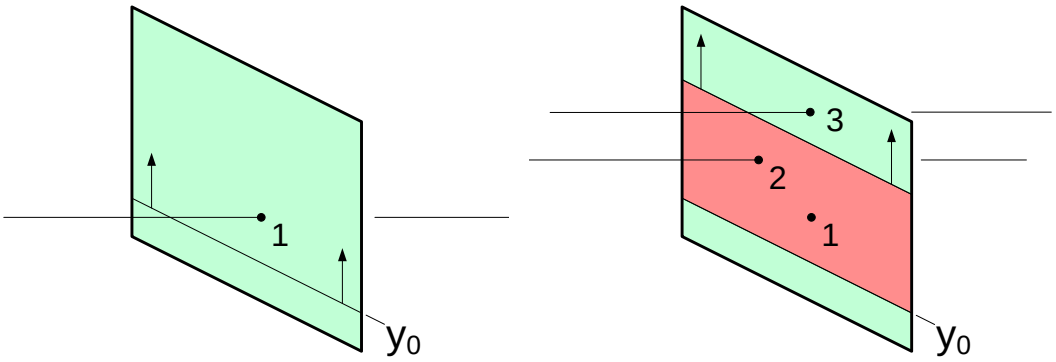


Figure 5.26: Efficiency for different cluster signal cuts. The black curve shows the efficiency observed without any restrictions. It turns out, that all lost hits are related to one of the three discussed system issues (first row masked, trigger latency and screen wiper effect). If these lost hits are not taken into account, an efficiency of 100 % is observed for cluster signal cuts lower than 1100 ADU.

Masked rows. Due to the first row effect discussed in chapter 4.4 in roughly 30% of the events the first two rows of double pixels which are read out show significantly higher signals than normal and are thus temporary masked. This issue can be avoided by a slower readout timing.

Trigger latency. The delay between a particle incident and the start of the readout after accepting the trigger lasts up to $\Delta t \approx 600ns$. Most of this delay is caused by trigger signal transition times of the cables and NIM modules, as the distance between the photo multipliers in the beam area, the NIM logic in the control room and back to the area was around 30 m. This already leads to a transition time of > 200 ns. Since the system was operated using the rolling shutter mode with a fast row rate of 5 MHz in the idle state (ready and waiting for a trigger) already three rows of double pixels are cleared within 600 ns. If the particle impact is located in these rows, the respective signals are lost. At the ILC this imposes no limit, as a continuous (triggerless) readout is foreseen. For further beam test studies this issue can be avoided by using a row rate during the idle state slower than the trigger latency. Since a higher row rate reduces the effects of leakage current, in addition a shorter signal path should be achieved by placing the discriminator and the coincidence logic closer to the system.

Sequential readout (screen wiper effect). The first two issues are only related to particles which already have passed the detector at the time the readout is started (in the following referred to as *initial* or trigger particles). In case of a high particle flux additional particles may pass the detector within the time needed to read out the sensor. Later particles will be detected, if their impact positions are located in a region of the matrix which has not been processed by that time (see figure 5.27). The probability to detect a later particle thus decreases linearly with the time of arrival. As the temporal



(a) (b)

Fig. 5.27: Illustration of the screen wiper effect. (a) Particle 1 triggers the readout starting with the following row (y_0). The readout direction is upwards, indicated by the arrows. (b) With a delay of Δt two additional particles pass the detector. Particle 2 hits a region which has already been processed and is therefore not detected. Particle 3 will be detected.

distribution of the particles during a spill is random, the average probability is 50%. There are different approaches to minimize the inefficiencies caused by this effect. One can reduce the hit multiplicity by either a faster readout (in particular using zero-suppression) or a lower particle flux. Another possibility is an extendable readout, i.e. as soon as an additional particle passes the detector during the readout the row counter is reset to assure that a full frame is read out after each particle.

Plausibility check. For a plausibility check all 308 missed hits are analyzed in respect to the three potential inefficiency issues. The resulting numbers are then compared with the expected numbers, which are calculated in the following.

First, the initial particles will be studied, which are missed due to the masked rows and the trigger latency. The number of predicted hits located in a masked row or within the last three rows read out is $n_{missed,\,1st\,particles} = 206$.
The expected number of hits lost due to these effects is calculated by:

$$n_{expected\,misses,\,1st\,particles} = n_{events} \cdot \frac{A_{ROI}}{A_{scintillator}} \cdot (f_{masked} + f_{latency}) \cdot \varepsilon_{tel} \tag{5.11}$$

where f_{masked} and $f_{latency}$ are the inefficient fractions of the sensor area. Three of the 64 rows of double pixels are inefficient due to the trigger latency ($f_{latency} = 3/64 \approx 4.7\%$) and $f_{masked} = 1.1\%$ of the gates are masked. The overlap region A_{ROI} of all detector planes corresponds to 11% of the scintillator area $A_{scintillator}$.
The DEPFET telescope efficiency ε_{tel} is also affected by these effects. Since track points are required in all four telescope planes the probability to find a track decreases to $\varepsilon_{tel} = [1 - (f_{masked} + f_{latency})]^4 = 79\%$. In addition, in three telescope planes several areas (roughly 5% of the overlap area) had to be masked due to significantly higher noise. Due to a different pixel layout one of the sensors (module 9) showed a signal to noise ratio

of only 20.5. The efficiency of this plane was found to be $\varepsilon_{mod9} \approx 54\%$. Due to these additional inefficiencies the final track efficiency is calculated to be $\varepsilon_{tel} \approx 35\%$.
Applying the stated values to equation (5.11) leads to an expected number of missed initial particles of $n_{expected\,misses,\,1st\,particles} \approx 201$ out of 90000 events, which is in good agreement with the observed number $n_{missed,\,1st\,particles} = 206$.
The question if a hit is missed due to the screen wiper effect is not easy to answer, as the exact time of the particle passage is not known. A track is only observed, if the particle has been detected in all four telescope planes, which in principal also suffer from the screen wiper effect. Therefore, the smallest distance of the track point to the row index at the time of the trigger y_0 (startgate)

$$\Delta_{tel,min} = y_{hit} - y_0 \tag{5.12}$$

found in one of the telescope planes can be used as an upper limit of the delay between trigger and particle passage. In figure 5.28 $\Delta_{DUT,missed}$, which is defined similarly as the distance between the predicted hit and the startgate of the DUT, is plotted versus $\Delta_{tel,min}$. In the case of $\Delta_{tel,min} < \Delta_{DUT,missed}$ the particle should have been detected in the DUT and a screen wiper effect can be excluded, whereas in the opposite case the screen wiper effect cannot be excluded.

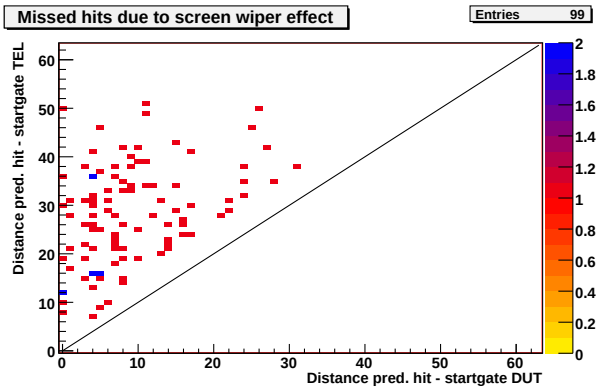


Figure 5.28: The closest distance between track point and startgate in one of the telescope planes is used as an upper limit of the particle delay relative to the trigger. If the distance of predicted hit and startgate in the DUT is smaller than in all telescope planes the screen wiper effect cannot be excluded.

To check if the observed number of missed particles is plausible, the probability for multiple events has to be approximated following [38]. The particle flux through the overlap region of the detectors A_D is given by:

$$\begin{aligned} \phi_D &= \int_x \int_y \phi(x,y)dxdy \quad d\tau \\ &= A_D \frac{dn}{dt} \quad d\tau \\ &= \frac{A_D}{A_T} \cdot \phi_T \end{aligned} \tag{5.13}$$

with a trigger area A_T and the frame readout duration $d\tau$. During the CERN beam test the raw trigger rate was measured to be $\phi_T \approx 1200\ ev/s$. The probability for multiple particles within a short timescale $d\tau$ after the trigger is given by the Poisson distribution:

$$p(n) = \frac{\phi_T^n}{n!} \cdot e^{-\phi_T} = \frac{\left(\frac{A_D}{A_T} \cdot \phi_T\right)}{n!} \cdot e^{-\frac{A_D}{A_T} \cdot \phi_T} \tag{5.14}$$

With a readout duration of a frame of $\tau_{frame} = 896\mu s$ and an overlap region of $A_D = 1.96mm^2$ the particle flux is estimated to be

$$\phi_{D,ROI} \approx 135ev/s \tag{5.15}$$

The probability of further particles passing the detector (i.e. a multiplicity of more than one) within the duration of a frame readout after the trigger is then given by

$$p_{mult} = 1 - e^{-\phi_D \cdot \tau} = 11.4\% \tag{5.16}$$

That this value is almost equal to the aspect ratio of scintillator and overlap region is a coincidence. In analogy to equation (5.11) the expected number of hits lost due to the screen wiper effect is calculated by:

$$n_{expected\,misses,\,2nd\,particles} = n_{events} \cdot p_{mult} \cdot \varepsilon_{tel,\,mult} \cdot 50\% \tag{5.17}$$

The telescope efficiency for multiple track events, i.e. the probability to find a hit in each telescope plane, is given by $\varepsilon_{tel,\,mult} = (0.5)^4 = 6.25\%$. Again, this efficiency is reduced by 59% due to the poor performance of a single detector plane and several masked areas in the other planes. This yields $n_{expected\,misses,\,2nd\,particles} \approx 130$ out of 90000 events, which is in agreement with the observed number $n_{missed,\,2nd\,particles} = 102$.

In conclusion, inefficiencies in the order of 10% were observed, which is much higher than expected. However, all missed hits could be assigned to one of the three sources of readout system related inefficiencies, i.e. the masked rows, the long trigger latency and the unsynchronized sequential readout. The observed numbers are in agreement with the expected numbers. The three discussed issues are due to the beam test setup and will not constitute a problem for the final vertex detector. However, inefficiencies that are unavoidable for the envisaged final readout scheme in the order of 0.1% for the final timing or 1.1% for the current timing could not be verified with the existing setup.

5.9 Detection purity

The detection purity is a measure of how many of the detected hits are related to real particles. It is defined as:

$$P = \frac{\text{no. of tracks}}{\text{no. of detected hits}} \tag{5.18}$$

where only tracks with a corresponding hit in the DUT are considered. Usually choosing a seed cut is a trade-off between efficiency and purity: the lower the cut the higher the efficiency; in turn, also the fraction of noise hits increases. Going to higher seed cuts, on the other hand, leads to a lower efficiency and a higher purity. This decision is particularly delicate for detectors with a low signal to noise. In experiments interested in b-physics a high purity is mandatory.
For a study of the purity in a beam test environment the reference system has to be perfectly efficient. If a detected hit in the DUT can not be assigned to a track due to an inefficient reference system a precise determination of the detection purity is not possible. Given the efficiency issues discussed in the last section a study of the purity is not reasonable in the present case.
However, the knowledge of the efficiency of the telescope planes allows an approximation of the expected purity. First, the number of expected tracks through the ROI is estimated:

$$\begin{aligned} n_{particles,\,expected} &= n_{initial\,particles} + n_{mult.\,particles} \\ n_{initial\,particles} &= n_{events} \cdot \frac{A_{ROI}}{A_{Scintillator}} = 9900 \\ n_{mult.\,particles} &= n_{events} \cdot p_{mult} = 10260 \end{aligned} \tag{5.19}$$

Then, the expected number of detected tracks with track points in all five planes (including the DUT) is estimated by:

$$\begin{aligned} n_{tracks,\,expected} &= n_{initial\,p.} \cdot \varepsilon_{tel,\,initial} \cdot \varepsilon_{DUT,\,initial} + n_{mult.\,p.} \cdot \varepsilon_{tel,\,mult} \cdot \varepsilon_{DUT,\,mult} \\ &= 9900 \cdot 0.35 \cdot 0.942 + 10260 \cdot 0.026 \cdot 0.5 \\ &\approx 3400 \end{aligned} \tag{5.20}$$

This number is in good agreement with the observed number of hits with a corresponding track of $n_{tracks,observed} = 3370$. The expected number of hits detected within the ROI of the DUT is calculated similarly:

$$\begin{aligned} n_{hits} &= n_{initial\,particles} \cdot \varepsilon_{DUT,\,initial} + n_{mult.\,particles} \cdot \varepsilon_{DUT,\,mult} \\ &= 9900 \cdot 0.942 + 10260 \cdot 0.5 \\ &\approx 14450 \end{aligned} \tag{5.21}$$

Based on these numbers the expected purity is $P_{expected} = 3400/14450 = 0.235$, which is in agreement with the observed purity of $P_{observed} = 0.205$. For this study, a seed cut of $6\sigma_{noisepeak}$ and a cluster cut of $1100ADU$ has been used. By lowering the cluster cut the presence of the low pulse height signals discussed in chapter 5.3 leads to a lower purity. These entries within the range of $6\sigma_{noisepeak}$ seed cut and the low signal tail of the Landau distribution, which make up roughly $\sim 10\%$ of the total amount of clusters, are not related to tracks.

5.10 Summary

Within the scope of this thesis the first ILC-type DEPFET beam test measurements have been prepared and carried out at beam facilities at DESY and CERN. DEPFET sensors of the PXD-4 production featuring 64×128 pixels with a pitch and $22 \times 36 \mu m^2$ were broadly tested. A continuous non-zero-suppressed data acquisition has been developed allowing a continuous readout rate of 250 Hz. Roughly 60% of the time is needed for the data transfer to the PC via the USB 2.0 link. With a frame rate of ~ 1.1 kHz the readout of the sensor is limited by the readout chip.
A signal to noise of $S/N > 120$ has been observed with a sensor of $450 \mu m$ thickness under beam test conditions without cooling. The cluster size distribution is dominated by 4 pixel clusters. Different reconstruction algorithms have been studied. The charge sharing particularly in the direction of the smaller pitch in conjunction with the η-reconstruction leads to an excellent spatial resolution of $\approx 2 \mu m$ including the error on the predicted position. These values will change for an envisaged sensor thickness of below $100 \mu m$. The signal will be correspondingly lower and also the charge cloud will be considerably smaller. Studies are needed as soon as a thinned sensor is available, which is planned for 2009.
For the first time a DEPFET beam telescope consisting of four PXD-4 sensors has been operated successfully, allowing for high precision beam test studies. Currently the in-pixel charge collection efficiency is under study.
Efficiency and purity studies were limited by systematic inefficiencies due to a long trigger delay and a high particle multiplicity. The missed particles could all be assigned to one of these issues, which will not constitute a problem for the final vertex detector.

6. Zero-suppressed readout

Within the scope of this thesis a DEPFET matrix has been successfully read out using the on-chip zero-suppression of CURO 2. The continuous data acquisition has been studied for different threshold settings with and without tuned thresholds under beam test conditions. A detailed analysis of the data revealed several deficiencies. The most severe are fluctuations of the supply voltage caused by the pre-charge mechanism leading to threshold instabilities. In particular a significant drop of the threshold in single rows has been observed. Furthermore, the timing used during the beam test lead to a general shift of the threshold to higher values. The analog baseline (pedestal offset) showed a dependence on the threshold setting, which had to be corrected off-line. In the following these issues are discussed and the methods for an appropriate event reconstruction are presented.
A tool to understand and verify the zero-suppressed data is to apply an off-line cut on non-zero-suppressed data discussed in the last chapter. Under the premise of a simulation of the on-chip zero-suppression only one cut is being applied for seed and neighbors before all off-line corrections of pedestals and common mode noise. Please note, that the on-chip threshold is applied after the second memory cell, the noise at the point of the hit discrimination is therefore expected to be smaller than with the simulation. The noise of the analog signal in turn is equal to the non-zero-suppressed case. Some of the results are compared to off-line zero-suppressed data.
Based on the outcome of the beam test studies an improved readout sequence has been developed and tested in the lab using a γ-source. Some of the results are presented in the following.

6.1 Timing issues

In a zero-suppressed readout scheme, the digital hit information is buffered on the chip, which can store up to 256 double hits. Since the I/O pins of the chip for the readout of the digital data are shared with other signals, the chip has to be configured for the readout via command and data register. For the serial readout the same clock signals are used as for the data acquisition, which requires a major break in the readout sequence. The implementation of the hitbuffer reset, which has to be applied at the end of the data transfer, foresees an additional WrCLK after the detection of the first hit. These constraints are particularly critical, as the WrCLK is involved to a high degree. Since the WrCLK steers the comparator pre-charge, fluctuations of the supply voltage after finding the first hit are likely.
As already discussed in chapter 4.4 the current design of the readout chip is very sensitive to the timing of the individual sampling steps. Compared to the non-zero-suppressed

readout the utilization of the comparator requires a modified timing.

Comparator pre-charge and sampling (*WrCLK*) While the high phase of the second WrCLK of a readout cycle (FIFO sample) has no special impact on the zero-suppression, during the low phase the pre-charge of the comparator is active.
By shorting the input and output node of the CMOS inverter, the input voltage is adjusted to the trip point. During this procedure the analog power consumption is highest, rising by $\sim 30\ mA$ for the whole chip. In a pulsed operation the stabilization of the resulting voltage drops of the analog supply takes some time. During the test beam the comparator pre-charge was set to 140 ns, which was found to be not sufficiently long enough.
In figure 6.1 a typical zero-suppressed event display acquired during the beam test is shown. The readout direction is upwards. The threshold was set to $\sim$ 8500 ADU. A particle hit consisting of four pixels is visible in the upper center. The row index where the trigger occurred and the readout started is not known. The first pixels above the threshold are the two pixels in row 111. The row of double pixels thereafter is completely read out. Note, that it is always the row after the first hit, independent of the startgate[1]. As the average signal in that row of $\sim$ 8000 ADU is a typical pedestal value, this is a different effect than the so-called *first row effect* discussed in chapter 4.4, where the analog baseline of the first row is shifted to higher values. Thus a common mode effect is excluded. In this case, the only explanation is a much lower threshold in the second read out row than in the others. This effect has been observed for lower threshold settings in numerous events, while with higher thresholds more and more pixels in the respective row are below the threshold and not read out. The difference of the two thresholds is constant and roughly 600 ADU. Given the amount of detected hits in the following rows, there is no evidence for significant threshold fluctuations in the following rows.

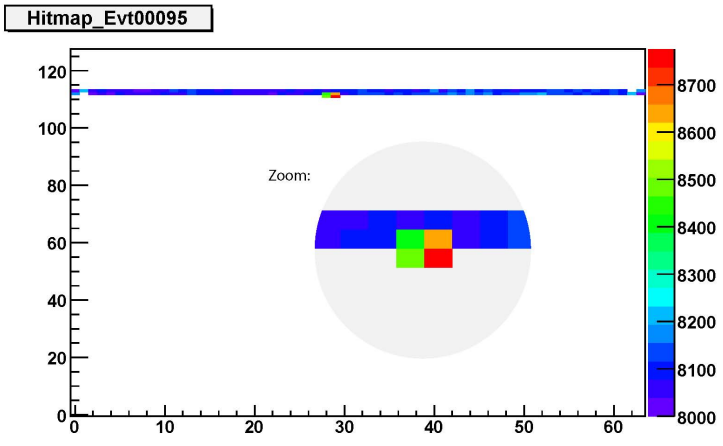


Figure 6.1: Typical event display acquired during the beam test. The readout direction is upwards. The threshold was set to $\sim$ 8500 ADU. A particle hit consisting of four pixels is visible in the upper center, spread over two rows of double pixels. The first row containing two hit pixels is read out normally. The following row of double pixels is completely read out due to a threshold drop.

After the beam test a readout sequence has been developed, which shows a significantly better performance, in particular no threshold fluctuations have been observed. Important modifications are the slower timing of the comparator pre-charge (380 ns) and the comparator sample duration (240 ns instead of 140 ns). The biggest improvement has been achieved by parking the WrCLK in the high state during the serial sampling of the

[1] See sec. 3.7 for a definition of the startgate.

analog values. The settling times of the input sampling and the sequential sampling of the analog values can be retained unchanged.

6.2 Event reconstruction

Since the noise hits are already suppressed on the readout chip, ideally no further cuts are needed in the off-line analysis. All contiguous signals are merged to clusters. After an offset correction discussed below, no further corrections are possible with the current generation of the readout chip. Given the high common mode values observed with a non-zero-suppressed readout, correction capabilities are necessary for the next chip iteration. This could be either an automatic on-chip correction or the readout of the mean row signal for an off-line correction.

6.2.1 Pedestal shift correction

With non-zero-suppressed analysis, a dedicated pixel-wise pedestal calculation is done. By subtracting the pedestals from the raw signal any kind of fixed offset is compensated. Provided the pedestals are not changing from run to run an off-line correction of zero-suppressed data is also possible. In the present case the pixel-wise pedestals vary from run to run and pedestal values are strongly dependent on readout timing, it is therefore not reasonable to transfer non-zero-suppressed values to zero-suppressed data sets. Furthermore, depending on the readout timing, pedestal baseline variations have been observed with different threshold settings. Based on the non-zero-suppressed sequence the pedestals were measured for different thresholds. Note, that in non-zero-suppressed readout the comparator is usually disconnected and the threshold value is set to 0. Connecting the comparator leads to a baseline shift of $\approx 70-80$ ADU. The results of a threshold scan using a non-zero-suppressed readout scheme are shown in fig. 6.2, demonstrating that the analog baseline is not independent from the threshold setting.
With zero-suppressed data the offset value can be determined by matching the peak values of the seed signals, which should be at the same analog value for different thresholds. If the threshold is set high enough to cut off parts of the seed signal, a convolution of gaussian (seed signal) and error function (threshold) is fitted to the data to assure a correct peak detection (see figure 6.3). For γ source measurements the maximum signal (high signal cut off) may be used as an indicator as well. The obtained values for an offset correction of the beam test data in terms of a matching seed signal are shown in fig. 6.2.

6.2.2 Threshold drop

Since the events with a fully read out row at low thresholds are too numerous to ignore, a work-around in the off-line analysis is needed. In case of more than 20 hits per row the average signal is used for a pedestal correction. A noise value of 50 ADU (1σ) is assumed, corresponding to a non-zero-suppressed value before all corrections. Finally, every signal larger than $3 \cdot \sigma_{noise}$ is accepted as a hit.

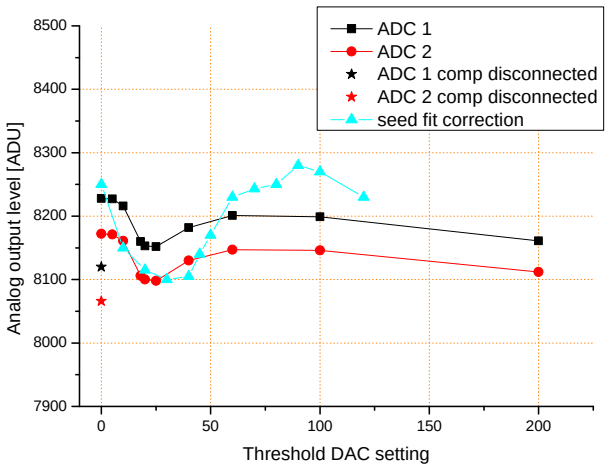
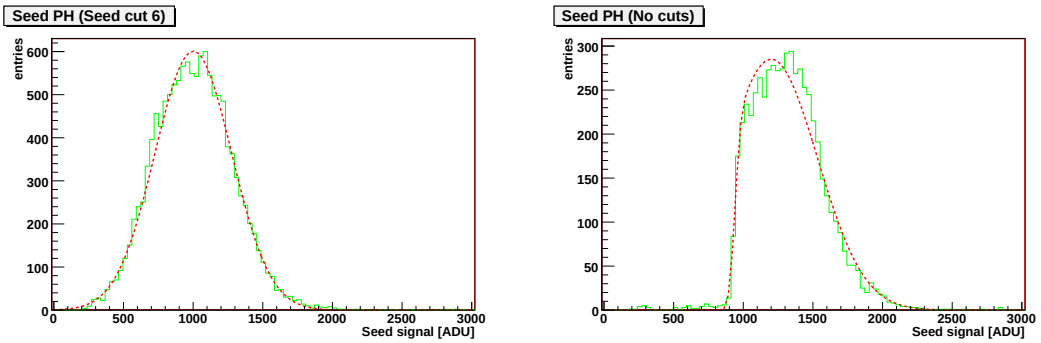


Figure 6.2: Mean pedestal value as a function of the threshold. The black (square) and red (circle) values have been measured with a non-zero-suppressed readout. The blue (triangle) curve has been determined by matching the zero-suppressed seed signals to the value obtained from a non-zero-suppressed readout.



(a) Non-zero-suppressed readout.

(b) Zero-suppressed readout.

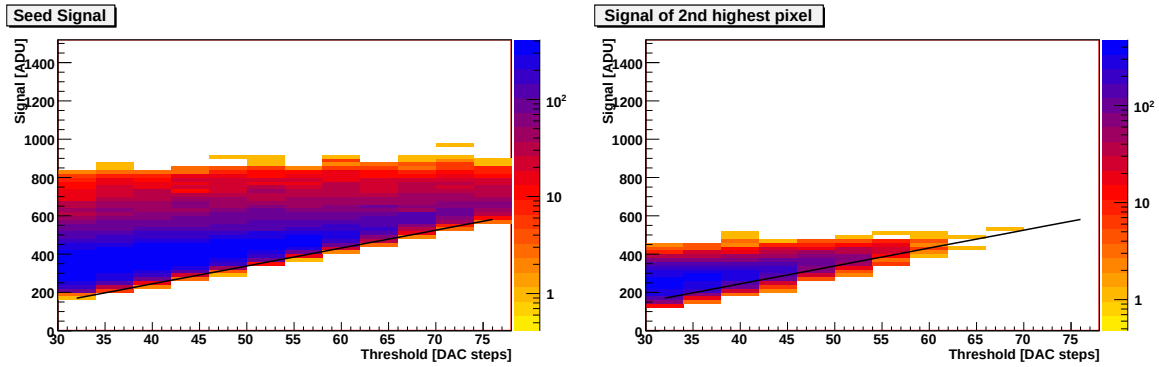
Fig. 6.3: Comparison of seed signals using non-zero-suppressed readout (left) and zero-suppressed readout (right). The lower seed signals are already below the threshold of 40 DAC units. For a correct peak detection a gaussian superimposed to an error function is fitted to the data.

6.3 Signal analysis

6.3.1 Threshold calibration

In fig. 6.4(a) the seed signal distributions are plotted for different threshold settings; the data is acquired using an Americium γ-source and the improved timing sequence. Fig. (b) shows the same plot for the second highest pixel. With increasing threshold the signals are truncated and the offset and gain of the threshold DAC can be extracted. The black line indicates the expected slope of the threshold DAC of 9.34 ADU or 232 e^- per DAC step (based on the calibration in chapter 4.5); the offset (threshold at a DAC value of 0) is adjusted to the data. The data is in good agreement to the expectations, the offset is found to be ≈ -130 ADU, which means that the threshold can be set well below the mean pedestal value.

For the calibration of the threshold offset and gain of the beam test data, the distributions



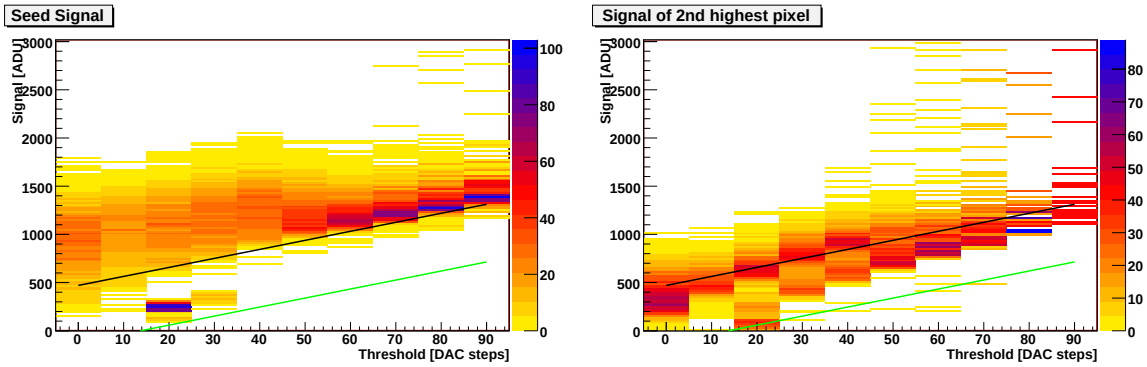
(a) Highest pixel (Seed). (b) 2nd highest pixel.

Fig. 6.4: Threshold calibration using optimized settings and an Americium γ-source. For each threshold the measured signal distribution is plotted in color coding. With increasing threshold the signal is cut off. The black line corresponds to the expected threshold according to the DAC calibration in chapter 4.5. The observed slope is in excellent agreement with the expectations. Note, that the optimized readout settings do not require a threshold dependent pedestal correction as the pedestal (and thus the high signal cut off) is found to be independent from the threshold.

are normalized in terms of signal entries. The results are shown in figure 6.5. The expected slope of the threshold is indicated by the green line, using the offset found with the calibration discussed above. It can be seen, that the data is in rough agreement to the expected slope, however, the offset is shifted to much higher values. Based on the black line adjusted to the data an offset of $\approx$ 12000 e^- (500 ADU) is observed. Thus, even with the lowest threshold no noise entries are observed, except for the completely read out rows with the threshold drops. Due to the threshold drops many entries below the threshold are observed, in particular the amount of entries of the second highest pixel below the threshold is noticeable. Since the data represents minimum ionizing particles there is no sharp high signal cut off as in the case of γ-particles. However, the differences are significant, which is another hint for unstable pedestal baselines due to a suboptimal timing.

6.3.2 Threshold scan

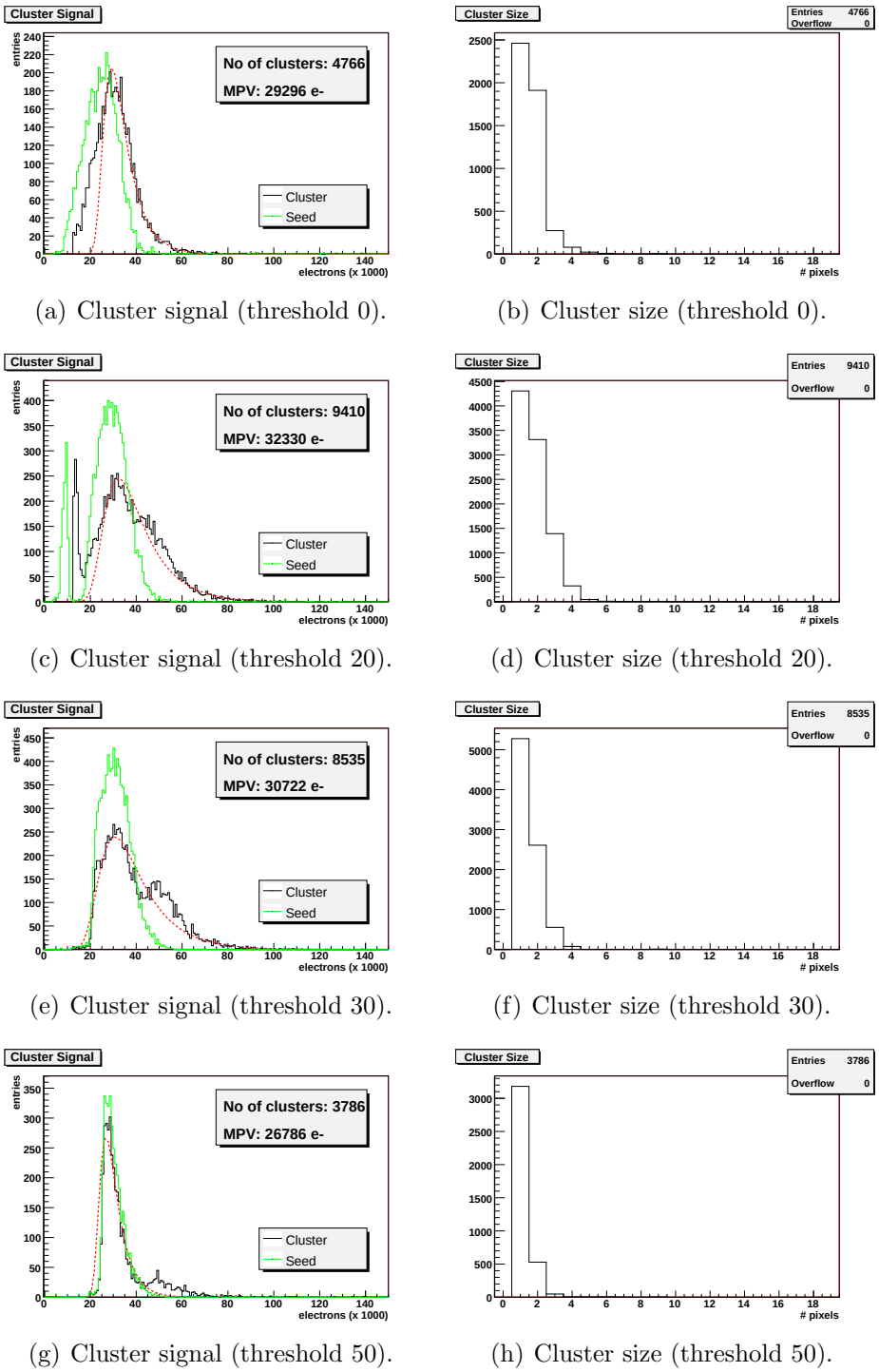
Figure 6.6 shows the cluster signal and the respective cluster size distributions for different thresholds based on beam test data. A most probable cluster signal using the lowest threshold (*threshDAC* setting of 0) of MPV = 28350 e^- has been observed. Compared to a non-zero-suppressed readout the cluster sizes are much smaller. Even with the lowest threshold the distribution is dominated by single hit clusters. Thus, even with the lowest threshold most neighbors carrying parts of the cluster signal are cut off, leading to a cluster signal which is $\approx$ 25% smaller than a non-zero-suppressed one. Due to the lack of off-line corrections for pixel-wise pedestals and row-wise common mode the signal distribution is wider.
With increasing threshold more and more signal is lost, leading to a pronounced separation



(a) Highest pixel (Seed). (b) 2nd highest pixel.

Fig. 6.5: Threshold calibration of the beam test data. The distributions are normalized in terms of signal entries. The slope of the green line corresponds to the expected threshold according to the DAC calibration in chapter 4.5. The experimental data of the seed signals (a) is in agreement with the expectations. Plot (b) shows the signal distributions of the second highest pixel, the entries below the threshold are significant. The threshold drops around 400-500 ADU.

of single and multi hit clusters. Also the seed signal is partly cut off. Finally, the detection rate decreases since many smaller clusters are below the threshold.
The results of a threshold scan using an Americium γ-source and the improved timing sequence are shown in figure 6.7. While with a threshold DAC setting of 16, corresponding to 480 e^-, noise hits are observed, with a setting of 17 (614 e^-) they are significantly suppressed. Provided that the noise peak is centered around zero, this would correspond to a noise figure of roughly 200 e^- at the comparator, which is in agreement with the expectations.
Due to the lower threshold the cluster sizes are larger compared to the beam test data, although the 59.5 keV signal is lower and the dimensions of the charge cloud are smaller. The cluster signal of $\sim 16000e^-$ is in good agreement with the non-zero-suppressed reference, thus no significant signal truncation is observed.



(a) Cluster signal (threshold 0).

(b) Cluster size (threshold 0).

(c) Cluster signal (threshold 20).

(d) Cluster size (threshold 20).

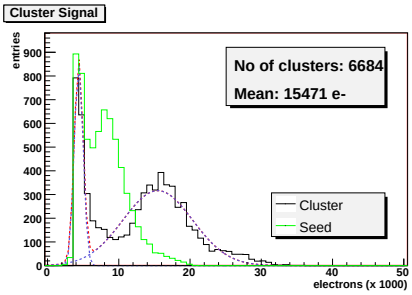
(e) Cluster signal (threshold 30).

(f) Cluster size (threshold 30).

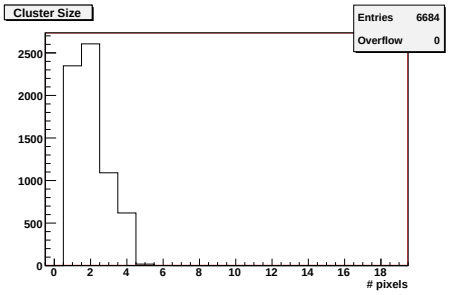
(g) Cluster signal (threshold 50).

(h) Cluster size (threshold 50).

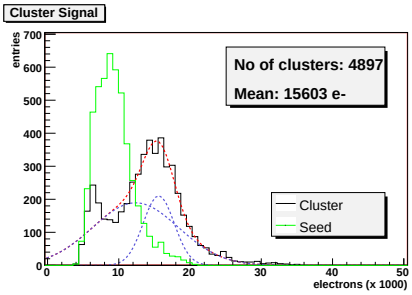
Fig. 6.6: Cluster signal and cluster size distributions for different thresholds. The green curve shows the seed signal, the black one shows the clustered signal. The data is based on beam test measurements. A convolution of a Landau distribution and an error function is fitted to the data with higher thresholds.



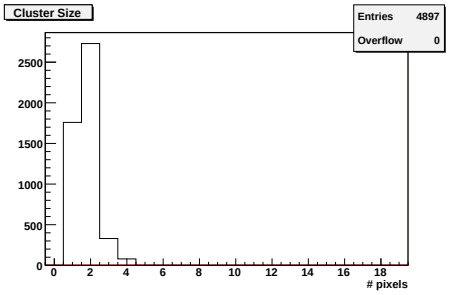
(a) Cluster signal (threshold 16).



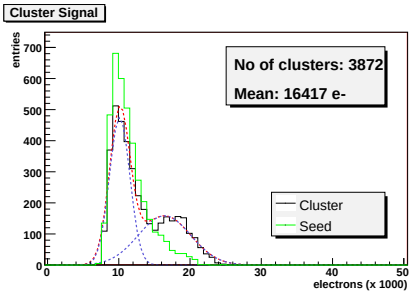
(b) Cluster size (threshold 16).



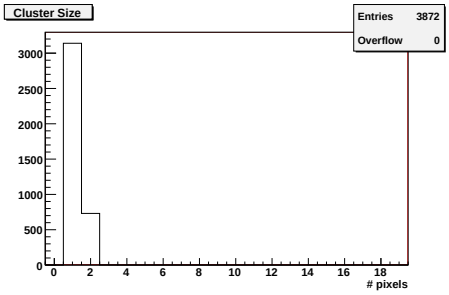
(c) Cluster signal (threshold 32).



(d) Cluster size (threshold 32).



(e) Cluster signal (threshold 60).

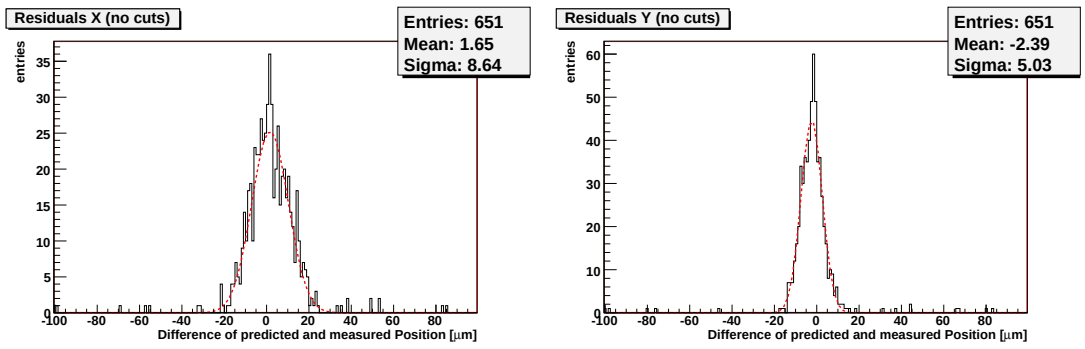


(f) Cluster size (threshold 60).

Fig. 6.7: Cluster signals for different thresholds using an improved timing sequence. A non-zero-suppressed ^{241}Am-spectrum is shown in figure 4.4 for comparison. The cluster signals with a peak value of $\sim 16000e^-$ are in good agreement, demonstrating that the zero-suppressed signal is fully reconstructed in case of a low threshold. It is possible to set the threshold just above the noise, allowing the reconstruction of events, which distribute charge on several pixels.

6.4 Spatial resolution

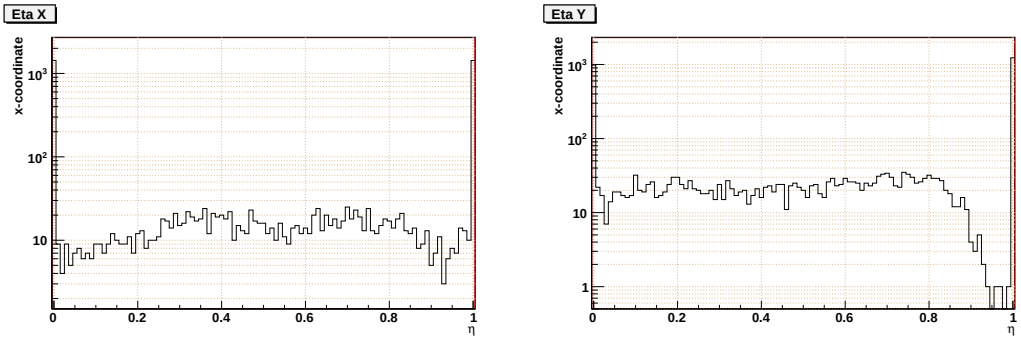
The spatial resolution of a zero-suppressed readout DEPFET sensor has been studied for different thresholds using different reconstruction algorithms. As shown in chapter 5.6.2 charge sharing between the pixels can be exploited for a significant improvement of the resolution. Applying a threshold on the pixel signals, particularly before the noise corrections, consequently results in a lower resolution. In the present case, the general shift of the threshold level to higher values leads to a cluster distribution, which is dominated by single hit clusters even at the lowest threshold.
The spatial reconstruction using the Center of Gravity method as defined in chapter 5.6.2 is possible without modifications. All pixels in a 3×3 area around the seed pixel are taken into account. The results are shown in figure 6.8 exemplarily for the lowest threshold of 0 DAC counts. Residuals of $(8.6 \pm 0.2)\ \mu m$ in x-direction and $(5.0 \pm 0.1)\ \mu m$ in y-direction are observed.



(a) x-direction. (b) y-direction.

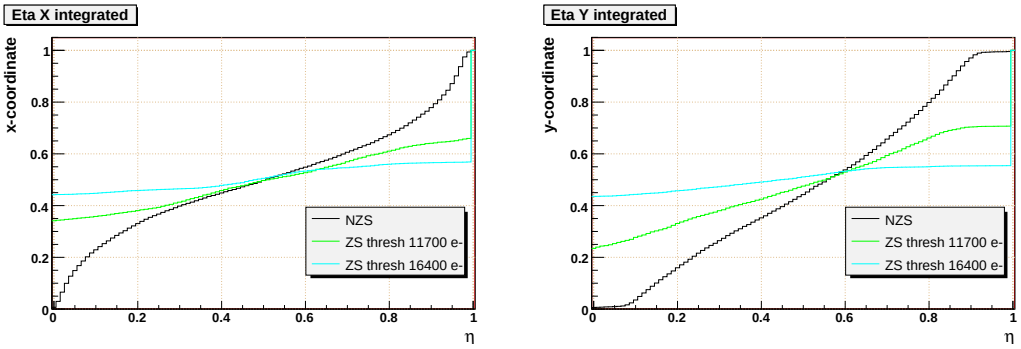
Fig. 6.8: Spatial residuals using the Center of Gravity method exemplarily for a threshold of 0 (untuned). Gaussian fits yield resolutions of $(8.6 \pm 0.2)\ \mu m$ in x-direction and $(5.0 \pm 0.1)\ \mu m$ in y-direction. The tracking uncertainties are still included. Note, that the DUT is one of the outer planes of the setup with a higher uncertainty of the predicted position.

The η-reconstruction as defined in equation (5.4) is not suited for zero-suppressed data without modifications, since always *two* signals in each direction are required. In case of no charge sharing and non-zero-suppressed readout the random noise determines which of the two seed neighbors is the second highest. This usually yields a symmetric η-distribution with equal entries close to 0 and 1. In the case of zero-suppressed data all signals below the threshold are exactly zero, leading to systematic preference of either 0 or 1. Therefore, the decision of assigning either value is made randomly by a random generator implemented in the analysis software. The resulting η-distributions for both directions are shown in figure 6.9. The final reconstruction distributions, i.e. the integrated η-distributions, for both directions are shown in figure 6.10. As shown in figure 6.11 exemplarily for the lowest threshold of 0 DAC counts the resolution using the η-reconstruction is found to be $(8.6 \pm 0.2)\ \mu m$ in x-direction and $(4.7 \pm 0.1)\ \mu m$ in y-direction.
These values still include the error on the predicted position. Since the zero-suppression



(a) x-direction. (b) y-direction.

Fig. 6.9: η-distributions for a threshold of 0. Please note the log-scale on the y-axis. The distribution is dominated by single pixel clusters, particularly in the x-direction.



(a) x-direction. (b) y-direction.

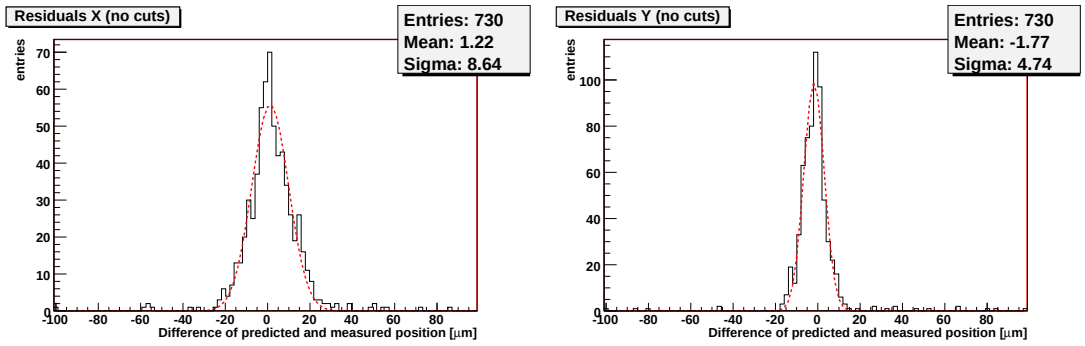
Fig. 6.10: Reconstruction functions $F(\eta)$ for both directions. For a comparison, the function $F(\eta)$ obtained with a non-zero-suppressed readout are shown. The areas close to 0 and close to 1 are more pronounced for higher thresholds as the probability for clusters without charge sharing increases.

is studied with one of the outer modules of the beam test setup, the error on the predicted position is relatively high. The simulation presented in chapter 5.7 is used for an approximation of this uncertainty. As the simulation assumes a perfect spatial reconstruction, the values of the non-zero-suppressed readout achieved with an η-reconstruction are used rather than the zero-suppressed ones. With a non-zero-suppressed readout of the module used for zero-suppression studies residuals of $(6.3 \pm 0.2)\ \mu m$ in the x-direction and $(3.6 \pm 0.1)\ \mu m$ in y-direction are observed, yielding the following tracking errors (see fig. 5.22):

$$\Delta x_{tel,TTTTD} = 5.0\ \mu m \tag{6.1}$$

$$\Delta y_{tel,TTTTD} = 2.8\ \mu m \tag{6.2}$$

Based on these values, the intrinsic resolutions achieved with a zero-suppressed readout at a threshold of $\approx 12000e^-$ are found to be:



(a) x-direction. (b) y-direction.

Fig. 6.11: Spatial residuals using the η-method exemplarily for a threshold of 0 and a seed cut of $6\sigma_{noise}$. The thresholds are not tuned. Gaussian fits yield resolutions of (8.6 ± 0.2) μm in x-direction and (4.7 ± 0.1) μm in y-direction. The tracking uncertainties are still included. Note, that the DUT is one of the outer planes of the setup with a higher uncertainty of the predicted position.

$$\Delta x_{ZS\,thresh\,0,\,TTTD} = 7.0\ \mu m \tag{6.3}$$

$$\Delta y_{ZS\,thresh\,0,\,TTTD} = 3.8\ \mu m \tag{6.4}$$

The observed resolutions as a function of the threshold are shown in figure 6.12. The values observed with the on-chip zero-suppression are compared to simulated values, obtained by applying an off-line threshold to the non-zero-suppressed data before all pedestal and common mode noise corrections. With the simulated data the best resolutions are obtained with a threshold between 5000 and 8000e^-. The charge sharing information is lost with higher cuts leading to a lower resolution. With a threshold of around 15000e^- and higher only events with very high or very low charge sharing are selected, effectively leading to a smaller pixel pitch. The higher resolution seen at high thresholds is therefore misleading. Thresholds higher than the most probable signal are not reasonable for the data acquisition. The studied threshold range is chosen for the comparison of simulated data and on-chip zero-suppression data.
Due to the threshold drop discussed above the resolutions obtained with the on-chip zero-suppression are lower than expected. The difference between the two reconstruction methods is not significant.

6.5 Threshold tuning

The spatial resolution has been studied in the beam test for both tuned and untuned thresholds. However, the difference between both data sets is found to be negligible. It turned out, that the tuning parameters were calculated using the uncalibrated tuneDAC leading to an over-compensation of the threshold dispersion. Further studies with a correct tuning are foreseen for the following beam test period.

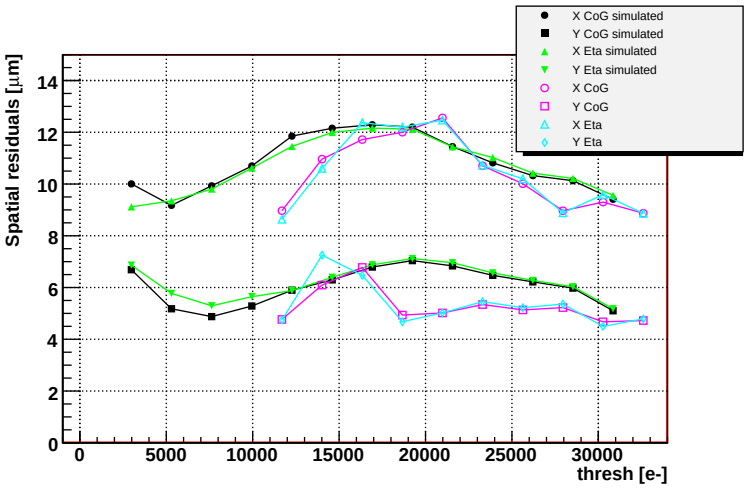


Figure 6.12: Spatial resolutions as a function of the threshold. The values obtained with the on-chip zero-suppression are compared to simulated values, obtained by applying an offline threshold to the non-zero-suppressed data.

6.6 Summary and Discussion

In this chapter the performance of the on-chip zero-suppression of the readout chip CURO 2 has been discussed based on measurements in a beam test environment and the laboratory.
The principal functionality of a continuous zero-suppressed readout of a DEPFET matrix with CURO 2 has been demonstrated. However, a detailed analysis of the beam test revealed several issues of the readout chip. Fluctuations on the supply voltage caused by a high power consumption during the comparator pre-charge lead to threshold fluctuations. Moreover, the timing used during the beam test lead to a general shift of the threshold, the lowest possible threshold with a DAC setting of 0 was found to be equivalent to $\approx 12000e^-$. That means, that a significant fraction of the cluster signal is below the threshold, compared to the non-zero-suppressed readout the cluster signal is $> 25\%$ lower. This also leads to a spatial resolution $\sim 30\%$ lower than with a fully analog readout. Due to the observed cluster size distribution, which is dominated by single pixel clusters, the difference between the studied reconstruction methods is negligible.
Due to a relatively high system noise and a wrong calibration a positive effect of the threshold tuning has not been observed.

It has been shown, that these design related issues can be compensated by a modified, particulary slower readout timing. The performance of the zero-suppressed data acquisition in the laboratory is encouraging. Unfortunately the modified sequence could not be tested in the beam test.
Based on these experiences a comprehensive re-design of the chip is necessary. The most important issue is the stabilization of the supply voltages. For a full evaluation of the triggerless readout envisaged for the ILC an analog FIFO and a modified serial protocol for the digital data is needed. An on-chip Common Mode correction would be beneficial and easy to implement.

7. Summary and Outlook

The planned International Linear Collider (ILC) requires a vertex detector of unprecedented precision to fully exploit the physics potential. One of the proposed technologies for the vertex detector is the DEPFET active pixel sensor. The proposal of the DEPFET collaboration foresees a silicon sensor of below $100\mu m$ thickness. The pixels of around $25 \times 25\mu m^2$ are read out row-wise. The expected occupancy and particulary the beam related background require a high row rate of up to 20 MHz. On-chip zero-suppression reduces the data volume considerably, but also the complexity of the readout chip increases. While for a zero-suppressed readout the pixel address of each hit is required, with a non-zero-suppressed scheme only the (analog) signals have to be transferred. The advantage of a zero-suppressed readout thus depends on the final occupancy.

With this thesis an ILC DEPFET prototype system has been developed, including the design of a new sensor, dedicated readout and steering chips, a DAQ PCB with a fast USB 2.0 interface, as well as data acquisition and analysis software. The first generation of an ILC-type DEPFET sensor of $450\mu m$ thickness features 64×128 rectangular pixels of $22 \times 36\mu m^2$. The concept of a fully current based, triggerless readout has been evaluated with the readout chip CURO 2, featuring on-chip pedestal subtraction and zero-suppression.

A detailed characterization of the system properties was performed in the laboratory using laser and γ-source measurements. The timing and operation parameters have been optimized. The internal gain of the sensor has been measured to be $g_Q = 308pA/e^-$. It has been shown, that a fast and complete clear is feasible using a voltage swing in the order of $13V$, facilitated by an unstructured deep-n implantation (highE) and a clear-gate structure. The system noise figure of $\approx 320e^-$ is dominated by the kT/C sampling noise contribution of the readout chip.

In the course of this thesis first ILC DEPFET beam test measurements were prepared and carried out at facilities at DESY and CERN. Most of the data is non-zero-suppressed, allowing comprehensive studies of system properties like signal, noise and charge sharing between the pixels. The most probable energy deposition of a minimum ionizing particle was found to be $(38250 \pm 50)\ e^-$, corresponding to a signal to noise ratio of $S/N = 121 \pm 3$. Charge sharing between the pixels yields typical cluster sizes of four pixels. The spatial reconstruction has been studied with different algorithms. The best results with residuals of $\sim 2\mu m$ in the direction of the small pitch have been achieved using the η-algorithm.

These values will change for an envisaged sensor thickness of below $100\mu m$. The signal will be correspondingly lower and also the charge cloud will be considerably smaller. Studies are needed once a thinned sensor is available, which is planned for 2009.

A DEPFET beam telescope has been developed and successfully operated allowing for high precision beam test studies (like in-pixel charge collection efficiency). Simulations yielded an uncertainty of the predicted position for the DUT placed in the center of the setup of below $2\mu m$.
For the first time a continuous readout of a DEPFET sensor using the on-chip zero-suppression was developed and studied under realistic conditions in a beam test. Substantial modifications of the system, the DAQ software and the analysis software were necessary to allow a continuous readout of analog and digital data.
For an evaluation of the envisaged ILC readout scheme the next readout chip of the CURO concept needs a full mixed signal FIFO, an adjustable coarse sampling and a modified digital readout protocol. With the present CURO 2 an average zero-suppressed row rate of ≈ 0.5 MHz has been achieved. Faster timing comes along with a signal degradation and threshold fluctuations. The spatial resolution has been found to be $\sim 30\%$ lower with a threshold of $12000e^-$ than with a fully analog readout.
Based on experience from a detailed beam test analysis, a modified readout sequence has been developed and tested in the lab. The performance is encouraging, issues like threshold fluctuations are not observed. Further studies in a beam test are needed for a comprehensive analysis.
The experiences particulary of the beam test studies are an invaluable input for the design of the next generations of the sensor and the readout chip. The concept of the double pixel structure and a static clear-gate of the sensor was found to be successful. The next iteration (PXD-5) has meanwhile been produced, featuring smaller pixels and potentially lower clear voltages. Lower clear voltages will allow the use of thinner gate oxides of the steering chips, yielding an improved radiation tolerance.
An alternative approach to the readout of the sensor is currently being studied with the DEPFET Current Digitizer (DCD). The signal current is immediately digitized after the pedestal subtraction by an 8-bit algorithmic current mode ADC for every channel. Since a zero-suppression is shifted to an off-chip processing a fast digital link of several 100 MHz is needed.

Bibliography

[1] Ankenbrandt et al. (Muon Collider Collaboration). Status of muon collider research and development and future plans. *Physical Review Special Topics - Accelerators and Beams*, 2, 1999.

[2] C. Sandow, L. Andricek, P. Fischer, R. Kohrs, H. Krüger, G. Lutz, H.-G. Moser, L. Reuen, R. Richter, L. Strüder, J. Treis, M. Trimpl, N. Wermes. Clear-performance of a linear DEPFET device. 2005.

[3] CALICE Collaboration. Report to the Calorimeter R&D Review Panel. Technical report, June 2007.

[4] DEPFET Collaboration. DEPFET Pixel Vertex Detector for the ILC. PRC report, 2007.

[5] DEPFET Collaboration. DEPFET Pixel Vertex Detector for the ILC (Report to the Vertex Detector R&D Review Panel). Technical report, 2007.

[6] Analog Devices. *AD8015 155 Mbps Transimpedance Amplifier Datasheet Rev. A*, 1996.

[7] Analog Devices. *AD9244 14-Bit 40/65 MSPS IF Sampling Analog-To-Digital Converter, Rev. C*, 2005.

[8] Zbynek Drasal. Geant4 testbeam simulations. Technical report, Charles University, Pague, Czech, 2007.

[9] J.J. Velthuis et al. A DEPFET based beam telescope with submicron precision capability. *Submitted to IEEE*, 2007.

[10] L. Andricek et al. The MOS-type DEPFET pixel sensor for the ILC environment. *Nucl. Instr. and Meth. A*, 565:165–171, 2006.

[11] R.H. Richter et al. Design and technology of DEPFET pixel sensors for linear collider applications. *Nucl. Instr. and Meth. A*, 511:250–256, 2003.

[12] W.-M.Yao et al. (Particle Data Group). Particle data booklet. J. Phys. G 33, 2007.

[13] E. Gatti and P. Rehak. Semiconductor Drift Chamber - an Application of a Novel Charge Transport Scheme. *Nucl. Instr. and Meth. A*, 225:608, 1984.

[14] Stephan Groneborn. Study of pair background due to beamstrahlung for different ILC beam parameter sets. *EUROTeV-Memo-2005_03_1*, 2005.

[15] Alan Hastings. *The Art of Analog Layout*. Tom Robbins, 2001.

[16] ICFA. International Linear Collider Reference Design Report. Technical report, February 2007.

[17] J. Treis, P. Fischer, O. Hälker, M. Harter, S. Herrmann, R. Kohrs, H. Krüger, P. Lechner, G. Lutz, I. Perić, M. Porro, R. Richter, L. Strüder, M. Trimpl, N. Wermes. Study of noise and spectroscopic performance of DEPMOSFET matrix prototypes for XEUS. *Nucl. Instr. and Meth. A*, 568:191–200, 2006.

[18] J. Ulrici, P. Fischer, P. Klein, G. Lutz, W. Neeser, R. Richter, L. Strüder, M. Trimpl, N. Wermes. Imaging Performance of a DEPFET Pixel Bioscope System in Tritium Autoradiography. *Nucl. Instr. and Meth. A*, 547:424–436, 2005.

[19] M. Karagounis. Entwicklung analoger integrierter schaltungen für pixeldetektoren. Master's thesis, Fernuniversität Hagen, 2004.

[20] J. Kemmer and G. Lutz. New semicondutor detector concepts. *Nucl. Instr. and Meth. A*, 253:356ff, 1987.

[21] M. Koch. CURO noise simulations. *internal note*, Rev. 1.0, 2007.

[22] M. Reiche R.H. Richter L. Andricek, G. Lutz. Processing of ultra-thin silicon sensors for future e+ e- linear collider experiments. *IEEE Transactions on Nuclear Science*, 51:1117 – 1120, 2004.

[23] L. Landau. *Journal of Physics 8*, 8:201, 1944.

[24] LDC Group. Detector Outline Document for the LDC. Technical report, 07 2006.

[25] W.R. Leo. *Techniques for Nuclear and Particle Physics Experiments*. 1987.

[26] G. Lutz. *Semiconductor Radiation Detectors*. Springer Verlag, Berlin, 1999.

[27] Markus Mathes. *in preparation*. PhD thesis, University of Bonn, 2008.

[28] Nicholas Walker (Editors) Nan Phinney, Nobukasu Toge. ILC reference design report. Technical report, 2007.

[29] Ivan Perić. *Design and Realisation of Integrated Circuits for the Readout of Pixel Sensors in High-Energy Physics and Biomedical Imaging*. PhD thesis, Physical Institute, Bonn University, 2004.

[30] R.-D. Heuer, D. Miller, F. Richard, P.M. Zerwas. TESLA Technical Design Report, Part III. Technical report, 2001.

[31] Lars Reuen. *in preparation*. PhD thesis, University of Bonn.

[32] Rainer Richer. private communication.

[33] Leonardo Rossi, Peter Fischer, Tilman Rohe, and Norbert Wermes. *Pixel Detectors*. Springer-Verlag, 2006.

[34] Christian Sandow. Ladungssammlung und löschvorgang von DEPFET-pixelsensoren. Master's thesis, University of Bonn, 2005. in german.

[35] Daniel Schulte. *Study of Electromagnetic and Hadronic Beackground in the Interaction Region of the TESLA Collider*. PhD thesis, Universität Hamburg, 1996.

[36] Helmuth Spieler. *Semiconductor Detector Systems*. Oxford Science Publications, 2006.

[37] R.D. Heuer R. Settles T. Behnke, S. Bertolucci. TESLA Technical Design Report, Part IV. Technical report, 2001.

[38] Johannes Treis. *Development and operation of a novel PC-based high speed beam telescope for particle tracking using double sided silicon microstrip detectors*. PhD thesis, Bonn University, 2002.

[39] Marcel Trimpl. *Design of a current based readout chip and development of a DEPFET pixel prototype system for the ILC vertex detector*. PhD thesis, Physics Institute, Bonn University, 2005.

[40] Adrian Vogel. Background Simulations for the International Linear Collider. 2007.